Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping

Vanessa Vermeirssen[1,4], Bart Deplancke[1,4], M Inmaculada Barrasa[1,4], John S Reece-Hoyes[1,4], H Efsun Arda[1], Christian A Grove[1], Natalia J Martinez[1], Reynaldo Sequerra[2], Lynn Doucette-Stamm[2], Michael R Brent[3] & Albertha J M Walhout[1]

Yeast one-hybrid (Y1H) assays provide a gene-centered method for the identification of interactions between gene promoters and regulatory transcription factors (TFs). To date, Y1H assays have involved library screens that are relatively expensive and laborious. We present two Y1H strategies that allow immediate prey identification: matrix assays that use an array of 755 individual *Caenorhabditis elegans* TFs, and smart-pool assays that use TF multiplexing. Both strategies simplify the Y1H pipeline and reduce the cost of protein-DNA interaction identification. We used a Steiner triple system (STS) to create smart pools of 4–25 TFs. Notably, we uniplexed a small number of highly connected TFs to allow efficient assay deconvolution. Both strategies outperform library screens in terms of coverage, confidence and throughput. These versatile strategies can be adapted both to TFs in other systems and, likely, to other biomolecules and assays as well.

Differential gene expression is regulated by TFs that bind to *cis*-regulatory DNA elements. To understand how differential gene expression is regulated at a systems level, it is important to systematically identify TF-DNA interactions and to model these interactions into transcription regulatory networks[1].

Gene-centered methods such as the Gateway-compatible Y1H system can be used for protein-DNA interaction (PDI) mapping for large sets of regulatory gene sequences[2]. So far, we have used this system to map two *C. elegans* PDI networks that contain PDIs for 72 and 38 gene promoters, respectively[3,4]. Both were mapped by screens with two prey libraries: a cDNA library[5] and a TF mini-library[2].

Comprehensive, genome-wide PDI mapping by Y1H screens would be relatively expensive and laborious because screens involve colony picking, processing and sequencing to determine the identities of interacting preys. Here we present two strategies that allow immediate prey identification, thereby streamlining the Y1H pipeline and reducing the cost. The 'matrix strategy' uses an array of individual TF clones (the 'TF array'), whereas the STS-based smart-pooling strategy (STS smart pooling) uses pools of TFs generated using an STS[6] and automated deconvolution of the assay readout. We found that Y1H matrix and STS smart-pooling assays both outperform library screens in terms of throughput, cost, confidence and coverage.

RESULTS

A *C. elegans* TF clone resource

Matrix or pooling Y1H experiments require a set of TF clones that is as comprehensive as possible. To create such a set, we first updated our *C. elegans* annotations from 934 to 940 TFs[7] (**Supplementary Methods** online). We obtained 571 TF-encoding open reading frames (ORFs) from the *C. elegans* ORFeome[8] and 30 from ongoing projects in our group[3,4] (data not shown). We reasoned that we could obtain some of the missing 339 TFs using updated gene models generated by TWINSCAN_EST[9] and/or GeneFinder (**Supplementary Fig. 1** and **Supplementary Table 1** online). For 212 TFs, we designed primers based on the shortest gene model that contains the DNA-binding domain. We obtained 154 TF-encoding ORFs (a 73% success rate), 117 of which were never cloned before. For 122 missing TFs, no new gene prediction was available. Altogether, we obtained ORF clones for 755 TFs (**Supplementary Table 1** and **Supplementary Fig. 1**), which includes members for the majority of TF families and represents 80% of all predicted *C. elegans* TFs. To our knowledge, this is the most comprehensive TF resource available to date. We transferred all TF ORFs into the Y1H prey vector $pDEST_{AD}$ by Gateway cloning[5,10] and refer to this collection as the 'TF array'.

Y1H matrix experiments

We first tested the TF array in Y1H matrix assays in which each bait is tested for interactions with each prey individually, and

[1]Program in Gene Function and Expression and Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. [2]Agencourt Bioscience Corporation, Beverly, Massachusetts 01915, USA. [3]Laboratory for Computational Genomics and Department of Computer Science and Engineering, Washington University, St. Louis, Missouri 63130, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to A.J.M.W. (marian.walhout@umassmed.edu).

RECEIVED 1 MARCH; ACCEPTED 23 MAY; PUBLISHED ONLINE 24 JUNE 2007; DOI:10.1038/NMETH1063

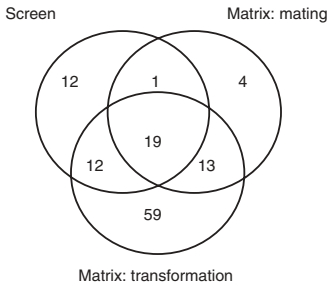

Figure 1 | Identified PDIs. Venn diagram of PDIs found by library screens (cDNA and TF mini-library combined) and the two matrix strategies (mating and transformation).

interacting preys are identified by their position in the array. Matrix assays were carried out both by transformation of TF plasmid DNA into haploid DNA bait strains, and by mating each DNA bait strain versus TF-transformed yeast strains. To allow efficient assay comparison, we selected four promoters that interact with many TFs as DNA baits (**Supplementary Table 2** online; an example of Y1H matrix results is shown in **Supplementary Fig. 2** online). We retrieved a total of 120 PDIs; 37 by mating and 103 by transformation, and recovered 32 of the 44 interactions found by library screens (73%; **Fig. 1**). Some interactions found only by library screens may not be recapitulated in matrix experiments because they can only be detected with truncated TF proteins that exclusively occur in cDNA libraries. Alternatively, not all TFs may function appropriately when they contain additional amino acids that are encoded by the Gateway vector tails flanking the TF ORF[11]. For instance, this may be the case for PAL-1 that was retrieved 14 times and only from the cDNA library with *Pdaf-3* (ref. 12; **Supplementary Table 2**). The coverage of matrix assays is higher than that of library screens as 76 new interactions were retrieved, 13 of which were retrieved both by mating and by transformation. One explanation for this is that in matrix experiments, a range of readout strengths that are above background can be detected (**Supplementary Fig. 2**).

Steiner triple system–based smart pooling

Y1H matrix assays carried out by transformation are effective and retrieve the largest number of PDIs, suggesting that their coverage is optimal. These assays, however, are still relatively laborious and expensive, as they require the transformation of 755 individual TF-encoding plasmids. Recently, several pooling strategies have been reported[13,14]. These can be referred to as 'smart-pooling' assays because they allow the immediate identification of positives by assay deconvolution, thereby avoiding retesting all components that are multiplexed in a positive pool[15]. An additional advantage of smart pooling is that each element is tested multiple (n) times in different pools, which provides an inherent experimental retest, and thereby reduces the number of false positive interactions[15]. The feasibility of smart-pooling assay deconvolution is related to the number of interactors that bind to a bait. For instance, when only one interactor binds a bait, n pools score positively, which is straightforward to deconvolute. When many interactors interact with a bait, however, deconvolution becomes increasingly difficult because more than one element in a pool may score positively. So far, the binary PI-deconvolution (pooling with imaginary tags followed by deconvolution) design is the only experimentally validated smart-pooling method[13]. The use of a PI-deconvolution design for *C. elegans* Y1H assays would require a 'batch size'[13] of 1,024 to accommodate all 755 cloned TFs. In such a design, each TF occurs in 10 pools (out of 20 pools in total), and pools comprise 512 TFs. Our data suggest that such a PI-deconvolution design cannot be used for Y1H assays because many TFs cannot be detected when they are diluted more than 27-fold (**Supplementary Fig. 3** online and data not shown). Moreover, such a PI-deconvolution design becomes impossible to solve for baits that interact with multiple TFs. We therefore reasoned that PI deconvolution could be feasible for Y1H assays when the design is partitioned into multiple smaller TF batches to accommodate a Y1H-compatible pool size. But modeling showed that such a design does not perform as well as our STS smart-pooling strategy.

Previously, we found that gene-centered PDI networks contain TF hubs[3,4]. Such hubs may hamper deconvolution when multiplexed with other TFs. Thus, our STS smart-pooling strategy uses a combination of 'uniplexing' and multiplexing (**Fig. 2a**,**b**). We separated the 755 cloned *C. elegans* TFs into two categories: we kept the 36 most frequently scoring TFs (with one or both Y1H reporters) separate, and multiplexed the remaining 719 TFs. We split these into three batches, each containing up to 256 TFs multiplexed into 52 pools (note that not all the 'real estate' of the design is used). This split has the added benefit of separating TFs that belong to a single TF family into different batches. Each multiplexed TF is present in three pools, each containing 4–25 TFs, which should allow detection of most PDIs by Y1H assays (**Supplementary Fig. 3**). Pairs of multiplexed TFs co-occur in a pool once at most. This facilitates the unambiguous deconvolution of Y1H positives (**Fig. 2c**). Our data indicate that 70% of promoters used in Y1H screens to date bind five or fewer TFs (**Fig. 2d**). Notably, this number increases to 88% when only multiplexed TFs are considered (**Fig. 2d**). This demonstrates that uniplexing highly connected TFs reduces the number of TFs requiring deconvolution.

We modeled the performance of the design *in silico* and compared it to a 'partitioned PI-deconvolution' design that has a pool size compatible with Y1H assays. The latter design partitioned 768 TFs into 12 batches. Each batch contained 12 pools, and thus the total design contained 144 pools. Each pool comprised 32 TFs, and each TF was present in six pools. We performed 1,000 simulations of both designs. As expected, we found that the unambiguous deconvolution of TFs is inversely related to the number of TFs that bind a promoter (**Fig. 2e**). We found that STS smart pooling outperforms PI deconvolution, as STS smart pooling identified 98% of all interactors unambiguously for promoters that bind five or fewer multiplexed TFs, compared to 73% for PI deconvolution (**Fig. 2e**).

The success of deconvolution is also affected by the false positive and negative rates of the assay used. A single pool that scores positively for a given interacting TF can either be the result of a single false positive pool or of two false negative pools. To ensure high-quality data sets, we defined a genuine interactor as a TF that scored positively in at least two pools. We modeled the STS smart-pooling design with levels of 10% and 20% of either false positives (**Fig. 2f**) or false negatives (**Fig. 2g**). As expected, the introduction of false positive or negative pools reduced the percentage of TFs that can be unambiguously identified. It is important to note that with additional sequencing all interacting TFs can be identified despite the occurrence of false positive pools. But not all TFs can be

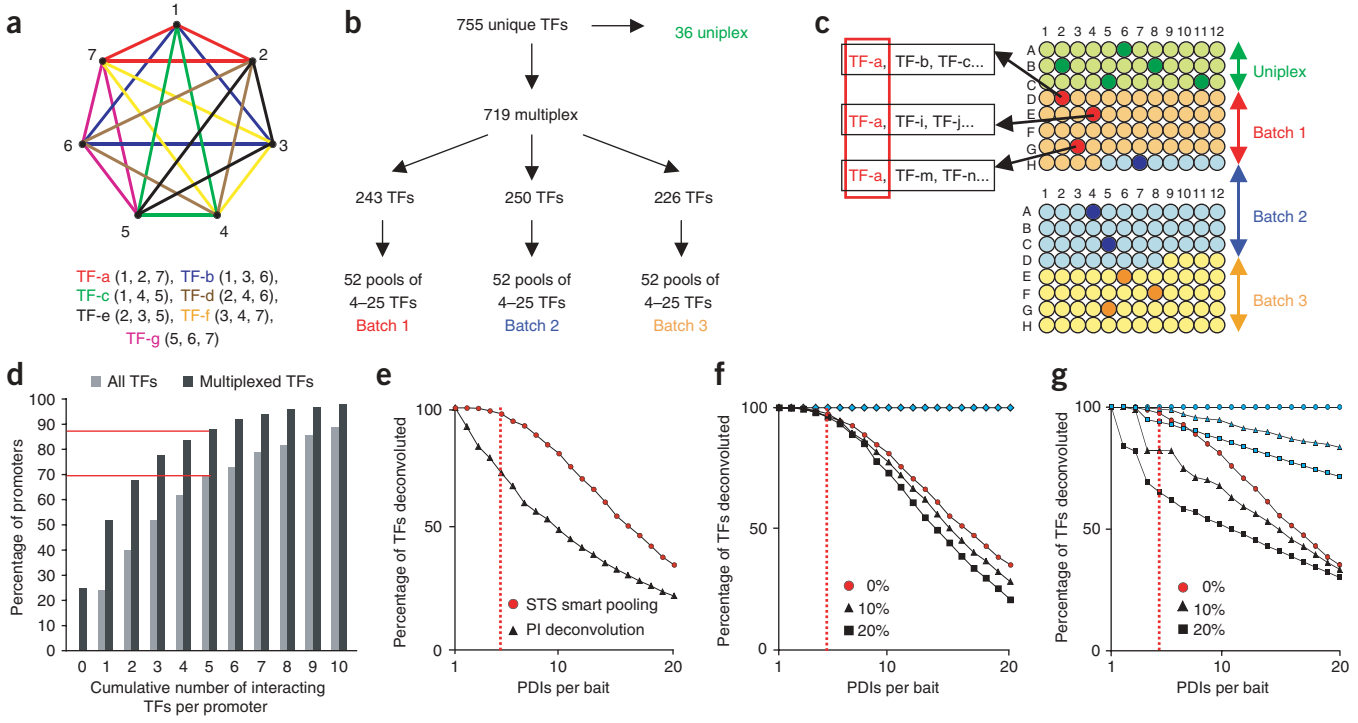

Figure 2 | STS-based smart-pooling design. (**a**) In an STS, S(2, 3, *v*) each element (here a colored triangle or a TF) is multiplexed in three pools (three values of *v*, 1–7 in the example) in such a way that each pair of TFs co-occurs once at most (that is, any two triangles in the figure have only one vertex in common). (**b**) Outline of the STS pooling design. The different colors indicate the 36 uniplex and three batches of multiplexed TF pools. (**c**) Visualization of a hypothetical result and deconvolution strategy; colors are as in **b**. (**d**) Most promoters bind five or fewer TFs (red lines). These data are based on Y1H results for more than 200 promoters. (**e**–**g**) Modeling of smart-pool design performance. The percentage unambiguously deconvoluted multiplexed TFs is plotted for different numbers of PDIs per bait. Comparison of STS pooling and PI deconvolution performance (**e**). Modeling STS pooling performance with false positive pools (**f**); with a limited amount of sequencing all interacting TFs can be recovered (blue diamonds). Modeling STS performance with false negative pools (**g**); blue symbols represent the corresponding percentages of TFs that can be recovered by sequencing.

identified when false negatives occur, even with additional sequencing, because sometimes two or even all three pools for a TF do not score. This is likely not a major problem because even with a 20% false negative rate, on average 94% of the interactors can be recovered for baits that bind five TFs (that is, by sequencing only six pools). Finally, all possible interactors can also be retested individually to discriminate false positives from false negatives.

Experimental validation of STS smart pools

We experimentally tested the performance of our STS smart-pooling design using Y1H assays by transformation into the four promoter baits that we also used in the matrix experiments. In total, we retrieved 98 PDIs, 30 of which were also found by library screens, and 74 of which were also found by matrix experiments (**Supplementary Table 2**). Of the 23 previously unidentified interactions, 10 involved uniplexed TFs and 13 involved multiplexed TFs.

The Y1H STS smart-pooling experiment using *Pfat-5* is shown in **Fig. 3** (a list of interactions is available in **Supplementary Table 3** online). For this bait, 11 uniplexed TFs and 20 TF smart pools conferred a positive Y1H readout. The deconvolution of 17 STS pools readily identified six TFs, all of which we confirmed by sequencing. CEH-6 could be deconvoluted unambiguously, even though only two of the three pools in which it occurs scored positively. We also determined the identity of the remaining interacting TFs by sequencing. Two pools contained TBX-9 and one contained CEH-18. CEH-18 was also retrieved in Y1H matrix experiments, suggesting that it is a genuine Y1H positive (**Supplementary Table 2**). Taken together, 24 pools should have given a positive readout (eight interactions times three, no overlap) and

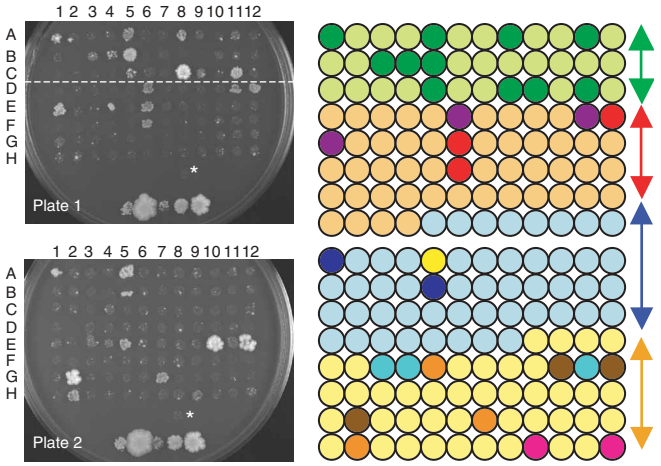

Figure 3 | Experimental validation of STS smart pools by Y1H assays using *Pfat-5* as a DNA bait. Interactors are indicated in the cartoon on the right and identities are provided in **Supplementary Table 3**. The spots at the bottom of each plate are Y1H controls. *, empty $pDEST_{AD}$ plasmid negative control.

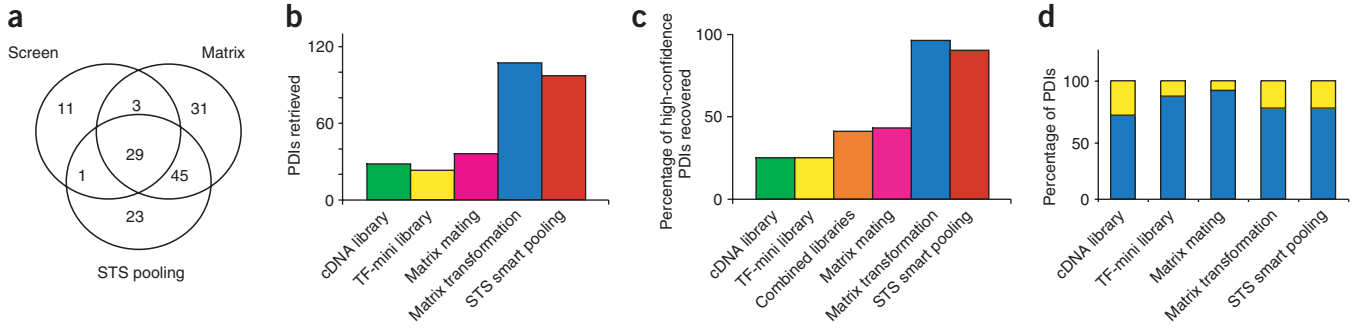

Figure 4 | Comparison of different Y1H assays. (**a**) Venn diagram of PDIs found by the three strategies. Screen: cDNA and TF mini-library combined; matrix: mating and transformation combined. (**b**) Total number of PDIs retrieved, separated by method. (**c**) Percentage of high-confidence interactions (that is, those found by multiple methods) retrieved with each method. (**d**) Proportion of high-confidence interactions per method (blue) or only found using the indicated method (yellow).

we detected 20. This indicates that four pools failed to confer a positive readout (17%, two for CEH-18 and one each for TBX-9 and CEH-6).

Comparison of Y1H methods

Next we compared the performance of the different Y1H methods (library screens, matrix experiments by mating and transformation, and STS smart-pooling assays; **Fig. 4a**). The absolute number of PDIs retrieved was highest for matrix transformation and STS smart pooling (**Fig. 4b**). To determine assay coverage, we only considered high-confidence PDIs, that is, those that were found by two or more Y1H methods (**Supplementary Table 2**). As expected, library screens provided the lowest coverage. Both the cDNA and the TF mini-library had ~25% coverage (**Fig. 4c**), but these were not completely overlapping as some interactions were found with one, but not the other library (**Supplementary Table 2**). Combined, the two libraries detected about 40% of the high-confidence PDIs (**Fig. 4c**). This level of coverage is similar to that of Y1H matrix experiments by mating. The highest coverage was obtained by transformation as the TF array and STS smart pooling both detected ~90% of the high-confidence interactions.

Each of the different methods also retrieved interactions that were not confirmed by another method. For both the transformed TF array and the STS smart pools, 23% of the interactions were exclusively found by that method (**Fig. 4d**), whereas 29% of the interactions found with the cDNA library were not confirmed by another method. We found that the interactions found by mating were most robust, as only 8% of these were not confirmed by another method. Taken together, the assay comparison data suggest that the coverage is highest for matrix transformation and STS pooling, whereas the confidence is optimal for matrix mating. A qualitative comparison of all the attributes of the different Y1H methods is available in **Table 1**.

Matrix and STS smart pooling in yeast two-hybrid assays

Yeast two-hybrid (Y2H) assays can be used to identify binary protein-protein interactions (PPIs), for instance, between dimerizing TFs[11]. As Y2H and Y1H assays use the same prey vector[17,18], we tested whether we could use the TF array and STS smart pools to detect TF dimers. As bait we used DB-NHR-49, because it is a hub in PPI networks[16]. We found that library screens and STS pooling detected 70% of the high-confidence interactions, whereas matrix assays detected 90% (**Supplementary Fig. 4** online). A similar separation of TF preys into uniplex and multiplex wells based on their overall connectivity properties for Y2H assays may further increase the number of PPIs that can be unambiguously deconvoluted using STS pooling.

DISCUSSION

The genome-wide analysis of transcription regulatory networks requires the comprehensive identification of TF-DNA and TF-TF interactions[1]. We present two new strategies, matrix and STS pooling assays, which can be used to find TF interactions with both Y1H and Y2H assays. Both strategies use an array of 755 TF clones.

The use of the TF array and STS smart pools provides major advantages over Y1H (or Y2H) library screens (**Table 1**). First, they provide much higher coverage, thereby reducing the number of false negative interactions in interactome data sets. The coverage is higher because each TF is tested directly, rather than recovered

Table 1 | Y1H and Y2H assay comparison

	cDNA library screen	TF library screen	TF array matrix by mating	TF array matrix by transformation	TF array STS smart pooling
Coverage	*	*	**	****	***
Confidence	**	***	****	***	***
Throughput: labor	*	*	***	**	****
Throughput: time	*	*	***	***	****
Cost	$$$$	$$$$	$	$$$	$$

Stars and dollar signs indicate quatitative strengths (stars) and cost (dollar signs), with a greater number of signs indicating an increase in that attribute.

from selected colonies. In library screens, many TFs will be missed because libraries are not screened to saturation. This is particularly likely for the *C. elegans* cDNA library, which is not normalized[5]. Additionally, TFs that interact weakly and provide only a small growth advantage are likely missed in library screens, which may explain why not all interactions are retrieved from TF mini-library screens. Second, the assays are much more cost-effective because they do not require extensive amounts of yeast colony PCR and sequencing. Sequencing is not required for matrix assays, whereas STS smart-pooling assays do require a small amount of sequencing to unambiguously identify all positive preys. In fact, our deconvolution program indicates which pools to sequence to both identify preys that cannot be unambiguously identified, and, if desired, to confirm the identity of preys that can. Even if all positive pools are sequenced, however, the number of sequencing reactions per bait is much lower than that of library screens. Third, the use of the TF array and STS smart pools not only allows the identification of positive preys, it also provides a set of negatives, that is, TFs that do not interact with a bait of interest. This will be helpful in determining the specificity of different TFs, for instance, within a particular TF family. This information will likely also be helpful for the determination of TF binding sites by computational methods. It is important to note, however, that cDNA library screens do have an advantage, in that they allow the identification of TFs that are not present in the TF array. These include predicted TFs for which a full-length ORF clone is not available, as well as putative TFs, that is, proteins that can interact with DNA baits but that do not have a recognizable DNA-binding domain[3]. In the future, we intend to include such putative TFs in our resources.

Y1H matrix assays can be carried out either by transformation or by mating. Each of these methods has advantages and disadvantages (**Table 1**). Y1H mating experiments are the least expensive, as they do not require the repeated purification of plasmid DNA. Mating experiments are also very robust, as 92% of the interactions found were confirmed by another Y1H method. Compared to transformation, however, the coverage of mating assays is relatively modest. Y1H transformation experiments provide the highest coverage, and the coverage of matrix assays is slightly higher than that of STS smart pooling. STS smart pooling provides several advantages over matrix transformations, including: (i) a fourfold reduction in cost for plasmid preparation (as two rather than eight 96-well plates are processed), (ii) a fourfold reduction in labor and (iii) an inherent retest as each multiplexed TF needs to score positively at least twice. By screening more than 200 *C. elegans* promoters, we found that 70% bind five TFs or fewer. Thus most promoters are suitable for STS smart-pooling assays. We did, however, occasionally observe promoter hubs, that is, promoters that bind a large number of TFs[4]. Such promoters may be less amenable to STS smart-pooling assays because when many STS smart pools score positively the deconvolution of the responsible TFs becomes problematic. It may therefore be desirable to reassay promoter hubs by matrix assays. It is important to note, however, that the promoters we used in this study all interact with a relatively large number of TFs, yet they performed well in STS smart-pooling assays.

STS smart pools combine multiplexing for the majority of TFs with uniplexing for highly connected TFs. This pooling strategy outperforms PI deconvolution[13]. Two other pooling strategies, 'two-phase mating'[19], and a shifted transversal design with uniform-sized pools[14] have also been described. The latter design performs well *in silico* (data not shown), but has yet to be experimentally validated. Two-phase mating can be used in Y2H assays, but is likely not as useful for Y1H assays. This is not only because of the large pool sizes that are used, but also because DNA baits interact with an average of five TFs, which results in a number of positives that precludes straightforward prey identification by this method. In other words, Y1H data are dense (an average of five positives out of 755 tested) compared to Y2H data (an average of five positives out of more than 10,000 tested)[11].

In summary, we developed a TF resource and two strategies to use it for PDI identification. An advantage of STS smart pooling is that it can be adjusted to suit different biomolecules and assays. For instance, our design may be adapted to Y2H assays by uniplexing PPI, rather than PDI hubs. Our STS pooling design is likely not only useful for mapping transcription regulatory networks using hybrid assays, but should be applicable to a variety of model systems and assays in which a medium-scale set of biomolecules needs to be analyzed.

METHODS

Creating the *C. elegans* TF-array. We obtained 571 TF ORF-containing Entry clones from the ORFeome (versions 1.1 and 3.1)[8,20] and transferred them into $pDEST_{AD}$ as described[5]. *Ab initio* TF ORF cloning was performed as described[5]. Updated gene predictions were generated either by GeneFinder (WormBase version WS145) or TWINSCAN_EST[9]. Primers were designed using OSP[21] (**Supplementary Table 1**). We obtained 30 TF ORFs from ongoing projects, eight of which were obtained from Y1H-positive yeast strains containing a cDNA library clone. These ORFs are not flanked by Gateway recombination sites (**Supplementary Table 1**). We verified all newly cloned TF ORFs by sequencing (Agencourt Bioscience Corporation). In total, the TF-array contains 757 clones representing 755 TFs (two isoforms are included for *daf-16* and *cdc-14*) distributed in eight 96-well plates. We transformed the TF fusions into Y1Hα001 (**Supplementary Methods**) as described[18]. The TF array will be available from Open Biosystems.

Steiner triple system–based smart-pool design and deconvolution. In a tertiary Steiner or STS of order v, S(2, 3, v), each element N (here a TF) is placed in three of a total of v pools in such a way that each pair of elements co-occurs at most once[6]. An STS exits (that is, has a solution) only if $v \bmod 6 = 1$ or 3, and the maximum number of elements N is given by $v(v-1)/6$. In theory, 69 pools (69 mod 6 = 3) can be used to accommodate all 755 TFs. In fact, a total of 782 TFs could be placed in such a design. But finding the solution to this design was not feasible, and we simplified it by (i) separating TFs into three batches and (ii) multiplexing a smaller number of TFs per batch than an optimal design would allow.

We aimed to use a maximum of two 96-well plates for yeast transformations. We decided to uniplex the 36 TFs that most frequently score in Y1H assays. This left 156 wells for pools of the remaining 719 multiplexed TFs. We split these wells into three sets of 52 pools (156/3 = 52). The closest optimal STS design that fills this space would use 51 pools (because 51 mod 6 = 3) each with the potential to accommodate 425 multiplexed TFs. But we used

an approximate design that allocates 256 TFs to 52 pools, accommodating a total of 768 TFs. In practice, we multiplexed 719 TFs, which corresponds to batches of 243, 250 and 226 TFs, and leaves space for an additional 49 TFs that may become available through ORF-cloning efforts. We developed a Java script that uses constraint programming to post the conditions of our design (that is, number of TFs, number of pools, maximum overlap of one), and the freely available constraint-solving package Choco (http://choco.sourceforge.net) to solve the problem design. A consequence of the search path followed by the solver is that pools contain an unequal number of TFs (4–25).

We performed the deconvolution with a Perl script that first determines which TFs are unambiguously identified. The script then selects one pool per TF for sequence confirmation for the unambiguously identified TFs, and a minimum set of pools to identify positives for which the corresponding TF cannot be unambiguously deconvoluted (that is, there are multiple solutions). A TF is unambiguously decoded if the three pools that contain the TF score positively and that TF is the only one that can account for the readout (that is, for all the other TFs in the pools the corresponding second and/or third pool are negative). The next round of deconvolution examines pools that did not have TFs for which all three pools score positively (that is, because of false negatives). In this case, a TF is unambiguously decoded if it is the only TF in that pool for which one of the other pools is also positive. An ambiguous positive occurs when multiple TFs in the pools can account for the positive readout. The STS design and deconvolution scripts are available from the EDGEdb database[12].

Additional methods. Descriptions of TF predictions, yeast assays and STS modeling are available in **Supplementary Methods**.

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

ACKNOWLEDGMENTS

We thank J. Dekker for critical reading of the manuscript; W. Harper (Harvard Medical School) for the Y187 yeast strain; N. Klitgord and M. Vidal (Dana-Farber Cancer Institute) for help with primer design; E. Méndez González (M-3 Informática, S.L.) for helpful suggestions regarding constraint programming and the sequencing staff at Agencourt Bioscience Corporation for technical assistance. This work was supported by BAEF fellowship for Biomedical and Biotechnology Research to V.V., and US National Institute of Diabetes and Digestive and Kidney Diseases grants (DK068429 and DK071713) to A.J.M.W.

AUTHOR CONTRIBUTIONS

V.V., B.D., J.S.R.-H, H.E.A., C.A.G. and N.J.M. performed all experiments; B.D., M.I.B. and A.J.M.W. conceived the pooling strategy; M.I.B. created the STS design and deconvolution, and performed the bioinformatics analyses. M.R.B. provided TWINSCAN predictions. R.S. and L.D.-S. provided sequencing; V.V., J.S.R.-H. and A.J.M.W. wrote the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods.

Published online at http://www.nature.com/naturemethods
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Walhout, A.J.M. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res.* **16**, 1445–1454 (2006).
2. Deplancke, B., Dupuy, D., Vidal, M. & Walhout, A.J.M.A. Gateway-compatible yeast one-hybrid system. *Genome Res.* **14**, 2093–2101 (2004).
3. Deplancke, B. *et al.* A gene-centered *C. elegans* protein-DNA interaction network. *Cell* **125**, 1193–1205 (2006).
4. Vermeirssen, V. *et al.* Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network. *Genome Res.;* published online 18 May 2007.
5. Walhout, A.J.M. *et al.* GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328**, 575–592 (2000).
6. Colbourn, C. & Mathon, R. in *The CRC Handbook of Combinatorial Designs.* (eds. Colbourn, C. & Dinitz, J.) 66–75 (CRC Press, Boca Raton, Florida, USA, 1996).
7. Reece-Hoyes, J.S. *et al.* A compendium of *C. elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol.* **6**, R110 (2005).
8. Lamesch, P. *et al. C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Res.* **14**, 2064–2069 (2004).
9. Wei, C. *et al.* Closing in on the *C. elegans* ORFeome by cloning TWINSCAN predictions. *Genome Res.* **15**, 577–582 (2005).
10. Hartley, J.L., Temple, G.F. & Brasch, M.A. DNA cloning using *in vitro* site-specific recombination. *Genome Res.* **10**, 1788–1795 (2000).
11. Walhout, A.J.M. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**, 116–122 (2000).
12. Barrasa, M.I., Vaglio, P., Cavasino, F., Jacotot, L. & Walhout, A.J.M. EDGEdb: a transcription factor-DNA interaction database for the analysis of *C. elegans* differential gene expression. *BMC Genomics* **8**, 21 (2007).
13. Jin, F. *et al.* A pooling-deconvolution strategy for biological network elucidation. *Nat. Methods* **3**, 183–189 (2006).
14. Thierry-Mieg, N. A new pooling strategy for high-throughput screening: the Shifted Transversal Design. *BMC Bioinformatics* **7**, 28 (2006).
15. Thierry-Mieg, N. Pooling in systems biology becomes smart. *Nat. Methods* **3**, 161–162 (2006).
16. Li, S. *et al.* A map of the interactome network of the metazoan *C elegans*. *Science* **303**, 540–543 (2004).
17. Deplancke, B., Vermeirssen, V., Arda, H.E., Martinez, N.J. & Walhout, A.J.M. Gateway-compatible yeast one-hybrid screens. *CSH Protocols* (doi:10.1101/pdb.prot4590; 2006).
18. Walhout, A.J.M. & Vidal, M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* **24**, 297–306 (2001).
19. Zhong, J., Zhang, H., Stanyon, C.A., Tromp, G. & Finley, R.L., Jr. A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating. *Genome Res.* **13**, 2691–2699 (2003).
20. Reboul, J. *et al. C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* **34**, 35–41 (2003).
21. Hillier, L. & Green, P. OSP: a computer program for choosing PCR and DNA sequencing primers. *PCR Methods Appl.* **1**, 124–128 (1991).