



Published in final edited form as:

*Nat Methods*. ; 8(12): 1050–1052. doi:10.1038/nmeth.1764.

## Yeast one-hybrid assays for gene-centered human gene regulatory network mapping

**John S Reece-Hoyes**<sup>1,2,3,\*</sup>, **A Rasim Barutcu**<sup>2,4</sup>, **Rachel Patton McCord**<sup>1,2,4</sup>, **Jun Seop Jeong**<sup>5,6,7,8,9,10</sup>, **Lizhi Jiang**<sup>5,6,7,8,9,10</sup>, **Andrew MacWilliams**<sup>11,12</sup>, **Xinping Yang**<sup>11,12</sup>, **Kourosh Salehi-Ashtiani**<sup>11,12</sup>, **David E Hill**<sup>11,12</sup>, **Seth Blackshaw**<sup>5,6,7,8,9,10</sup>, **Heng Zhu**<sup>5,6,7,8,9,10</sup>, **Job Dekker**<sup>1,2,4,11</sup>, and **Albertha J M Walhout**<sup>1,2,3,11,\*</sup>

<sup>1</sup>Program in Systems Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA

<sup>2</sup>Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, MA 01605, USA

<sup>3</sup>Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA

<sup>4</sup>Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

<sup>5</sup>Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>6</sup>Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>7</sup>Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>8</sup>Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>9</sup>Center for High-Throughput Biology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>10</sup>Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

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](mailto:marian.walhout@umassmed.edu)) or J.S.R-H. ([john.reece-hoyes@umassmed.edu](mailto:john.reece-hoyes@umassmed.edu)).

### AUTHOR CONTRIBUTIONS

A.J.M.W. conceived the project; J.D., J.S.R-H and A.J.M.W. designed the project; J.S.R-H and A.R.B. performed the experiments; J.S.R-H. and A.J.M.W. analyzed the data; J.J., L.J. and A.M. cherry-picked transcription factor ORF clones; J.J., L.J. assisted A.R.B. and J.S.R.-H. with Gateway cloning. R.P.M. performed the binding site analysis; H.Z., S.B., K.S.A., X.Y., A.M. and D.E.H. provided transcription factor ORFeome clones; J.S.R-H and A.J.M.W. wrote the manuscript.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

<sup>11</sup>Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

<sup>12</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

## Abstract

Gateway-compatible yeast one-hybrid (Y1H) assays provide a convenient gene-centered (DNA-to-protein) approach to identify the repertoire of transcription factors that can bind a DNA sequence of interest. We present a set of Y1H resources, including clones for 988 of 1,434 (69%) predicted human transcription factors, for the interrogation of interactions using either low or high-throughput settings. These approaches detect both known and novel interactions between human DNA regions and transcription factors.

---

Interactions between regulatory genomic DNA and transcription factors provide the first level of gene control and, therefore, the backbone of gene regulatory networks. Two complementary types of approaches can be used to identify such interactions<sup>1</sup>. Transcription factor-centered, or protein-to-DNA approaches (*e.g.* chromatin immunoprecipitation or ChIP) identify genomic regions bound by a transcription factor. Gene-centered, or DNA-to-protein approaches (*e.g.* Y1H), on the other hand, define the repertoire of transcription factors that can bind a DNA fragment of interest. The advantages and disadvantages of these techniques are discussed elsewhere<sup>2-4</sup>. In Y1H assays, a target DNA sequence (“DNA bait”) is cloned upstream of two reporter genes (*HIS3* and *LacZ*) to generate two DNA bait::reporter constructs<sup>5</sup>. After integration of these constructs into the yeast genome to generate a “DNA bait strain”, interacting transcription factors (“protein preys”) can be identified either by screening complex cDNA or transcription factor mini-libraries, or by testing individual protein preys in a directed pair-wise manner<sup>5,6</sup>. Activation of the *HIS3* reporter permits growth on media lacking histidine and containing 3-Amino-1,2,4-Triazole (3AT), a competitive inhibitor of the His3 enzyme, while activation of the *LacZ* reporter is detected by a colorimetric assay in which beta-galactosidase turns X-gal into a blue compound.

We have previously combined Y1H assays with Gateway cloning to transfer multiple DNA baits in parallel into the two Y1H reporter Destination vectors<sup>5</sup> and have applied these assays to delineate *C. elegans* gene regulatory networks<sup>7-10</sup> and to screen *Arabidopsis* gene promoters<sup>11</sup>. In an accompanying paper, we describe the development of a novel *C. elegans* Y1H pipeline, referred to as “enhanced Y1H” or eY1H<sup>4</sup>. eY1H employs a robotic setup together with an arrayed collection of yeast strains expressing transcription factor preys that can be mated with a DNA bait strain.

Currently, no gene-centered assays are available to identify human DNA-transcription factor interactions in a high-throughput manner. Here, we present a resource of human transcription factor-encoding open reading frames (ORFs) fused to the Gal4 activation domain (Gal4-AD) and apply this collection to several Y1H configurations, including the high-throughput eY1H pipeline, to map human regulatory interactions.

The human genome encodes 1,434 regulatory transcription factors, 1,116 of which are currently available in large clone collections<sup>12,13</sup> (**Methods** and Supplementary Table 1). We transferred these ORFs to the Gal4-AD Y1H prey vector by Gateway cloning and, after sequence verification, obtained 988 full-length transcription factor prey clones (Fig. 1a and Supplementary Table 1). These clones can be transformed directly into DNA bait strains for haploid-based Y1H experiments<sup>6</sup>. We transformed these clones into the Y1H prey strain to generate a human transcription factor yeast array (Supplementary Table 2) that can be used in small-scale mating-based Y1H experiments as well as in eY1H assays<sup>4</sup>. We also added 236 clones for unconventional DNA-binding proteins (uDBPs)<sup>13</sup> (Supplementary Tables 2 and 3).

To test the use of Y1H assays for the identification of human interactions, we first generated a small positive reference set (PRS) by literature curation (Supplementary Table 4). We predominantly focused on the well-studied beta-globin locus, but also included a few other regulatory regions and gene promoters (Supplementary Table 5). We tested each of the PRS interactions in different types of Y1H assays: in haploids and diploids, at different readout times, and under different Y1H conditions, because it has been demonstrated in other yeast-based assays that varying assay format and conditions results in a more comprehensive dataset<sup>14</sup>. Altogether, we could detect 24 out of 31 known interactions (77%, Fig. 1b,c and Supplementary Table 4), with five of these interactions detected in eY1H assays (16%)(Fig. 1d indicates one such interaction between the HBG1 promoter and NFYA). While this PRS detection rate for eY1H assays is comparable to that observed for high-throughput yeast two-hybrid screens<sup>15</sup>, this result contrasts to that observed with *C. elegans* bait strains where eY1H is at least as sensitive as the other Y1H methodologies<sup>4</sup>. This disparity is likely due to the fact the set of DNA baits used in this study exhibit higher levels of background as compared to most *C. elegans* baits<sup>4</sup> as well as other human DNA baits we have examined (data not shown). In fact, all of the interactions not found by eY1H were detected with only the *HIS3* reporter gene because of the high levels of background *LacZ* expression (data not shown). Essentially, the performance of the various Y1H approaches is intrinsically linked to bait strain behavior, and for optimal results the user can modulate the experimental settings accordingly (Supplementary Fig. 1).

With human eY1H assays, in which we screened 14 baits against the entire collection, we detected 175 DNA-protein interactions involving 13 DNA baits and 100 proteins (Supplemental Table 6). The proteins detected include 95 human transcription factors (~10% of the 988 tested) and five uDBPs (~2% of the 236 tested)<sup>13</sup>. The eY1H interactions do not exhibit a major bias for or against a particular type of DNA-binding domain (Fig. 2a), complementing our observations in *C. elegans* experiments<sup>7</sup>. We did find a larger proportion of nuclear hormone receptors, however, with most of these exclusively interacting with the CSF1 promoter (Supplementary Table 6), suggesting this enrichment is likely due to the small sample size of DNA baits.

Validating DNA-transcription factor interactions in complex systems is challenging and a “true negative” is nearly impossible to demonstrate<sup>3</sup>. While several of the interactions detected by eY1H are part of the PRS, and thus are known to have *in vivo* relevance, we wanted to assess the overall quality of the eY1H dataset. To this end, we evaluated the

relationship between transcription factor interactions observed in eY1H assays and their reported DNA binding sites. We first compiled DNA binding specificity information of human transcription factors or their orthologs<sup>16</sup> (**Methods**). Based on the potential transcription factor binding sites present within each DNA bait, we predicted which factors were expected to bind. Subsequently, we compared our experimentally detected factors to these predictions. An area under the receiver operating characteristic curve was calculated for each DNA bait sequence (Fig. 2b and Supplementary Fig. 2). We found significant enrichment for transcription factor binding sites in five of 11 DNA baits (Supplementary Table 7). This compares favorably to a similar analysis for ChIP-seq data (Supplementary Table 8 and 9), which suggests that eY1H assays may be more likely to capture direct physical interactions between DNA and transcription factors than ChIP. Two DNA baits (the MPL and HBB promoters) do not exhibit a correlation between the predicted transcription factor binding sites they harbor and the transcription factors retrieved in eY1H assays. This could be because these factors require interactions with co-factors and so are missed in eY1H, or because of differences in binding sites between human transcription factors and their orthologs. However, a more likely explanation is the high background reporter gene expression we observed for both these DNA baits, which makes them difficult to assay.

We visualized all eY1H interactions using Cytoscape<sup>17</sup>, generating the first gene-centered human gene regulatory network (Fig. 2c and Supplementary Fig. 3). Although small in size, we can already observe both specific as well as more promiscuous transcription factors. For instance, four factors each bind five DNA baits (*i.e.*  $k_{out} = 5$ ), while the majority of factors only interact with a single DNA bait (*i.e.*  $k_{out} = 1$ ). We also find several instances where multiple members of a transcription factor family can interact with the same DNA sequence. For instance, we found all four nuclear factor 1 transcription factors (NFIA, NFIB, NFIC and NFIX) interacting with the MPL promoter. This observation could reflect that these transcription factors have similar DNA binding specificities, and may be relevant in different cells or tissues, or under varying physiological conditions.

In summary, we have developed a collection of human transcription factors prey clones and a human transcription factor yeast array and combined these resources with our newly-developed eY1H platform<sup>4</sup>, facilitating the mapping of human gene-centered regulatory networks. The human eY1H pipeline will provide a powerful complement to transcription factor-centered methods, by enabling large-scale characterization of the DNA-binding activity of transcription factors that may be expressed or active only under restricted conditions or in a few cells. However, these resources can also easily be used for mating or direct DNA transformations of one or a few human DNA baits in small-scale studies.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

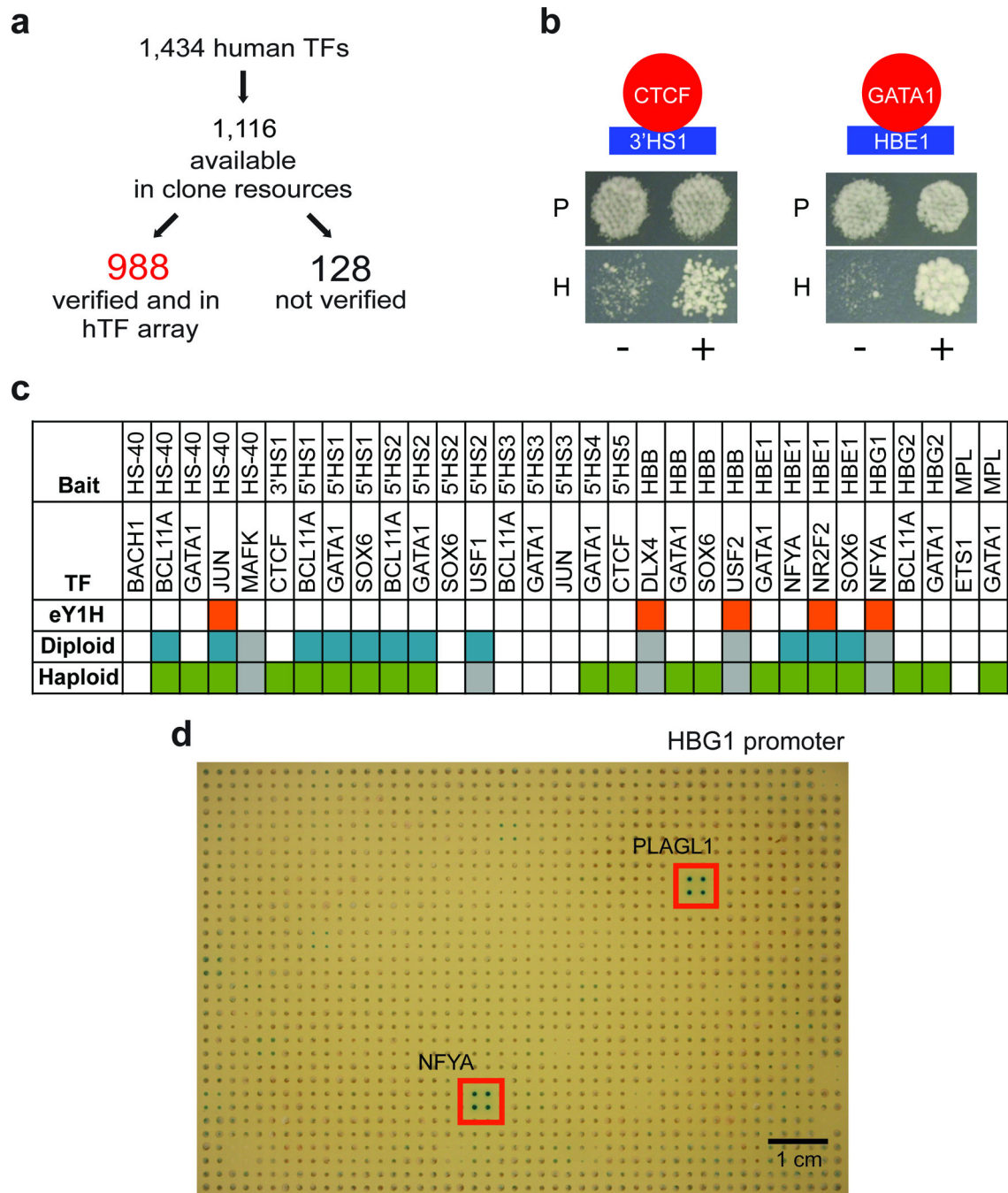
We thank members of the Walhout lab for discussions and critical reading of the manuscript. Research in the Walhout lab is supported by National Institutes of Health (NIH) grants DK068429 and GM082971. This work is supported by NIH grant HG003143 and a W.M. Keck Foundation Distinguished Young scholar award, awarded to

J.D.; an Ellison Foundation (Boston, MA) grant and Dana Farber Cancer Institute Sponsored Research funds awarded to Center for Cancer Systems Biology; Research in the Blackshaw lab is funded by a W.M. Keck Foundation Distinguished Young scholar award and a grant from the Ruth and Milton Steinbach Fund, H.Z. is supported by NIH grant GM076102.

## 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
3. Walhout AJM. What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biol.* 2011; 12:109. [PubMed: 21489330]
4. Reece-Hoyes JS, et al. Enhanced yeast one-hybrid (eY1H) assays for highthroughput gene-centered regulatory network mapping. *Nature Methods.* 2011 In press.
5. Deplancke B, Dupuy D, Vidal M, Walhout AJM. Gateway-compatible yeast one-hybrid system. *Genome Res.* 2004; 14:2093–2101. [PubMed: 15489331]
6. 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]
7. Deplancke B, et al. A gene-centered *C. elegans* protein-DNA interaction network. *Cell.* 2006; 125:1193–1205. [PubMed: 16777607]
8. 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]
9. 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]
10. Arda HE, et al. Functional modularity of nuclear hormone receptors in a *C. elegans* gene regulatory network. *Mol. Syst. Biol.* 2010; 6:367. [PubMed: 20461074]
11. Brady S, et al. A stele-enriched gene regulatory network in the Arabidopsis root. *Mol. Syst. Biol.* 2011; 7:459. [PubMed: 21245844]
12. Lamesch P, et al. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics.* 2007; 89:307–315. [PubMed: 17207965]
13. 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]
14. Chen YC, Rajagopala SV, Stellberger T, Uetz P. Exhaustive benchmarking of the yeast two-hybrid system. *Nat Methods.* 2010; 7:667–668. [PubMed: 20805792]
15. Braun P, et al. An experimentally derived confidence score for binary proteinprotein interactions. *Nat. Methods.* 2009; 6:91–97. [PubMed: 19060903]
16. Badis G, et al. Diversity and complexity in DNA recognition by transcription factors. *Science.* 2009; 324:1720–1723. [PubMed: 19443739]
17. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498–2504. [PubMed: 14597658]
18. 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]
19. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* 2009; 10:252–263. [PubMed: 19274049]
20. Fulton DL, et al. TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol.* 2009; 10: R29. [PubMed: 19284633]
21. 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.
22. Deplancke B, Vermeirssen V, Arda HE, Martinez NJ, Walhout AJM. Gateway-compatible yeast one-hybrid screens. *CSH Protocols.* 2006

23. Newburger DE, Bulyk ML. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* 2009; 37:D77–D82. [PubMed: 18842628]
24. Bryne JC, et al. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* 2008; 36:D102–D106. [PubMed: 18006571]
25. consortium, E.p. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007; 447:799–816. [PubMed: 17571346]
26. Granek JA, Clarke ND. Explicit equilibrium modeling of transcription factor binding and gene regulation. *Genome Biol.* 2005; 6:R87. [PubMed: 16207358]



**Figure 1.** Human gene-centered Y1H assays. **(a)** Schematic showing generation of the human transcription factor collection. The 1,116 available human transcription factors were transferred to the AD-2 $\mu$  Y1H prey Destination vector and 988 of the resulting clones were sequence-verified. **(b)** Examples of the detection of two known interactions in Y1H assays. Only *HIS3* activation is shown as the DNA baits exhibit high background levels of *LacZ* expression (data not shown). P – permissive growth media; H – media for detecting *HIS3* activation. **(c)** Detection of PRS interactions in different Y1H configurations. Orange –

eY1H; teal – diploids by mating; green – haploids by transformation; grey – not tested. **(d)**  
Example of eY1H readout plate with the HBG1 promoter as DNA bait. Positive interactions are indicated.

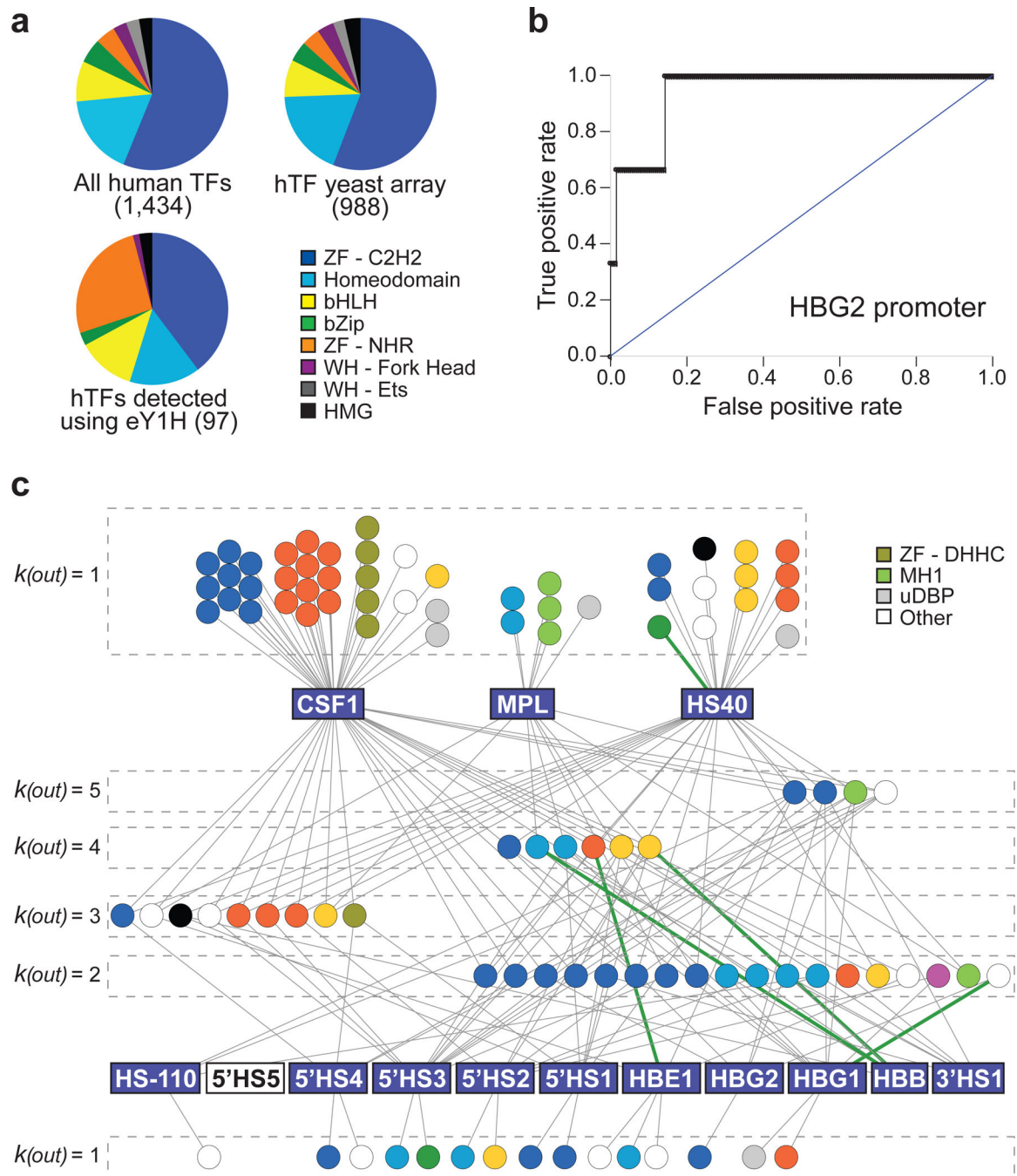
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Figure 2.** eY1H data analysis. **(a)** DNA binding domain analysis of the transcription factor compendium (1,434 in total), the transcription factor prey yeast array (988 in total), and the transcription factors detected in Y1H interactions (97 in total). ZF – zinc finger; bHLH – basic region helix-loop-helix; ZF - NHR – nuclear hormone receptor; WH – winged helix; HMG – high mobility group. **(b)** ROC curve of DNA binding site analysis for the HBG2 promoter. Binding sites within the DNA bait sequence are ranked from best to worst (in terms of match to PWM) along the x-axis – only the “best” binding site match for each

transcription factor is used. As the curve progresses along the x-axis, it steps up only for binding sites of transcription factors detected by eY1H. If the binding sites provided no information regarding eY1H interactions (*i.e.* no match between interactions predicted and detected), the curve would be largely below the diagonal. **(c)** Transcription factor-DNA interactions detected by eY1H depicted in a gene regulatory network (see Supplementary Figure 3 for a more detailed view). The DNA bait 5'HS5 is depicted as a clear box because it had no interactions. Unless otherwise noted, colors indicate transcription factor families as in (a). MH1 - MAD homology 1 domain. The out-going degree  $k(out)$  (*i.e.* number of DNA baits bound per transcription factor) is indicated. Green edges indicate detected PRS interactions.