

In vitro analysis of DNA–protein interactions by proximity ligation

Sigrun M. Gustafsdottir*, Joerg Schlingemann, Alvaro Rada-Iglesias, Edith Schallmeiner, Masood Kamali-Moghaddam, Claes Wadelius, and Ulf Landegren

Rudbeck Laboratory, Department of Genetics and Pathology, Uppsala University, Dag Hammarskjöldsväg 20, SE-75185 Uppsala, Sweden

Communicated by Charles R. Cantor, Sequenom, Inc., San Diego, CA, December 17, 2006 (received for review August 10, 2006)

Protein-binding DNA sequence elements encode a variety of regulated functions of genomes. Information about such elements is currently in a state of rapid growth, but improved methods are required to characterize the sequence specificity of DNA-binding proteins. We have established an *in vitro* method for specific and sensitive solution-phase analysis of interactions between proteins and nucleic acids in nuclear extracts, based on the proximity ligation assay. The reagent consumption is very low, and the excellent sensitivity of the assay enables analysis of as few as 1–10 cells. We show that our results are highly reproducible, quantitative, and in good agreement with both EMSA and predictions obtained by using a motif finding software. This assay can be a valuable tool to characterize in-depth the sequence specificity of DNA-binding proteins and to evaluate effects of polymorphisms in known transcription factor binding sites.

ChIP | EMSA | proximity ligation assay

Sequence-specific DNA binding by proteins controls processes such as replication, recombination, and transcription. Distinct DNA-binding proteins are also associated with centromeres, telomeres, and other specialized regions in the genome, where they regulate chromosome condensation, cohesion, and other aspects of genome maintenance. Activation of genes by DNA–protein interactions is a fundamental regulatory mechanism involving the recruitment of chromatin-modifying complexes and transcription complexes to initiate RNA synthesis (1). Of all human genes, 6% are estimated to code for DNA-binding transcription factors (2). Alterations of DNA-binding proteins are frequently involved in human disease, particularly in cancer, but information about the sequence specificity of DNA-binding proteins remains limited.

Genome-wide location maps have been generated for binding sites for most transcription factors in yeast by combining ChIP (3) with DNA microarray analysis (ChIP-chip) (4–6). Other studies have focused on chromatin remodeling factors, such as components of the RCS complex (7) and the DNA replication machinery, e.g., ORC and MCMs (8). To identify human transcription factor binding sites, promoter or CpG island microarrays have been used (9, 10). Moreover, high-resolution tiling path arrays have been used for transcription factor binding studies (11). In this regard, the ENCODE (Encyclopedia of DNA Elements) project was launched in 2003 to identify all functional elements in the human genome (12), assisted by the development of PCR-product and oligonucleotide tiling path arrays initially covering 1% of the genome. Despite its power and utility, ChIP-chip analysis has to be considered a screening technology of moderate resolution.

Several other array-based techniques have been developed to identify genomic binding sites of transcriptional activators or other DNA-binding proteins (13–15), but in all cases results must be confirmed by independent reference methods. EMSA (16) is frequently used to verify binding specificity *in vitro*. The supershift assay is an improvement of the technique (17) that allows antibody-based identification of DNA-binding proteins in the gel-retardation assays. EMSA allows resolution of complexes of different stoichiometry or conformation, and it can identify DNA-binding proteins and their binding se-

quences within a gene's regulatory region. EMSA may also be used for quantitative analysis of thermodynamic and kinetic parameters. However, the method is associated with several drawbacks because it is time-consuming, requires large amounts of input material, and involves the use of both acrylamide and radioactivity. Furthermore, EMSA is difficult to adapt to high-throughput analysis.

Here, we present a procedure for measuring DNA–protein interactions that is based on the proximity ligation assay (PLA) for ultrasensitive protein analysis (18, 19). In this technique, specific DNA representations of detected proteins are created and amplified in place of the proteins themselves, analogously to the detection of mRNA as cDNAs via reverse-transcription reactions. Briefly, oligonucleotides are attached to specific protein-binding reagents, typically mono- or polyclonal antibodies. When two proximity probes recognize and bind the same target molecule or a complex of two interacting target molecules, the ends of their conjugated oligonucleotides come sufficiently close to allow them to be joined by enzymatic ligation, assisted by the addition of a connector oligonucleotide. The detected protein molecules thus promote the ligation reactions by ensuring sufficient proximity between the ends of the proximity probes' oligonucleotide extensions. As two independent recognition events are required for detection of the target protein to create an amplifiable DNA molecule, there is a conceptual similarity to PCR, ensuring highly specific detection. The ligation products can then be replicated, detected, and quantified by real-time PCR, whereas unreacted probes remain silent. In this fashion, complicated and insensitive detection of protein molecules is replaced by a straightforward, sensitive, and specific detection reaction for nucleic acids. The conversion of target molecule identities to DNA reporter sequences offers the additional advantage that large sets of proteins could be analyzed in parallel, e.g., by including distinct tag sequences in the amplified segments and subsequent readout by hybridization to oligonucleotide arrays (20). To use PLA for sensitive and specific detection of interactions between proteins and nucleic acids, one of the proximity probes is a partly double-stranded oligonucleotide with a single-stranded 3' extension. The other probe is an antibody directed against the DNA-binding protein, and it has an attached DNA strand with a free 5' end (Fig. 1). We demonstrate the benefits of this new assay by characterizing the DNA

Author contributions: S.M.G., M.K.-M., and U.L. designed research; S.M.G., J.S., and A.R.-I. performed research; A.R.-I. and E.S. contributed new reagents/analytic tools; S.M.G., J.S., and A.R.-I. analyzed data; and S.M.G., J.S., A.R.-I., C.W., and U.L. wrote the paper.

Conflict of interest statement: U.L. is the inventor of patents describing the proximity ligation technology. He is one of the founders of the company Olink AB, which exploits the proximity ligation technology.

Abbreviations: PLA, proximity ligation assay; S/N, signal-to-noise ratio.

*To whom correspondence should be addressed. E-mail: sigrun.gustafsdottir@genpat.uu.se.

This article contains supporting information online at www.pnas.org/cgi/content/full/0611229104/DC1.

© 2007 by The National Academy of Sciences of the USA

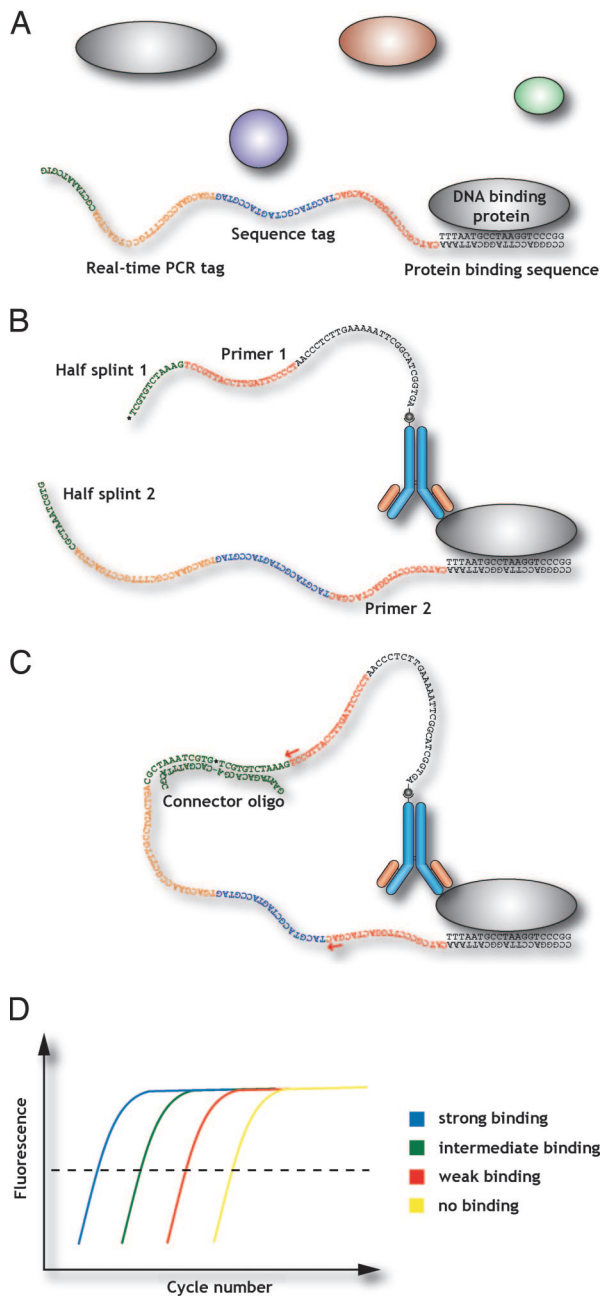


Fig. 1. Analysis of DNA-protein interactions by proximity ligation. A DNA-binding protein can be investigated by using two affinity probes, each carrying an oligonucleotide extension. One of the probes would be an antibody directed against a particular DNA-binding protein, whereas the other would consist of a partially double-stranded DNA sequence potentially recognized by the same DNA-binding protein. If the protein were simultaneously bound by both affinity probes (A and B), the ends of their appended oligonucleotides would be brought sufficiently close so that they could hybridize together to a subsequently added connector oligonucleotide, allowing them to be joined by enzymatic ligation (C). The ligated DNA sequence, which would serve as a specific DNA representation of the binding event between the protein and the investigated recognition sequence, would be subsequently amplified and detected by real-time PCR (D).

sequence specificity of the proteins p53, HNF-4 α , and USF1.

Results

We chose to analyze the transcription factors p53, HNF-4 α , and USF1 because they have been intensely studied and because they have all been associated with human disease.

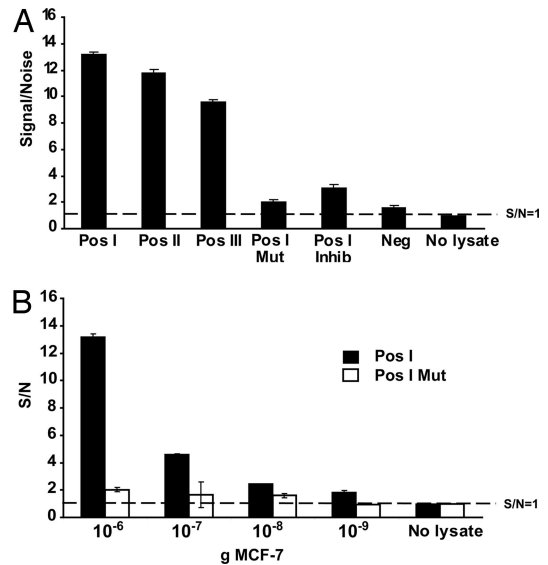


Fig. 2. Analysis of the DNA binding specificity of p53. (A) Three sequences previously reported to bind p53 were analyzed by PLA along with specificity controls in nuclear lysates prepared from MCF-7 cells. Positive controls: Pos I was a p53 consensus sequence published by Kastan *et al.* (23), Pos II (24) was a focal adhesion kinase promoter region shown to interact with p53, and Pos III was a polymorphic microsatellite that mediates induction of p53-inducible gene 3 by p53 (25). Negative controls: Pos I mut was a probe in which four nucleotides within the consensus binding site had been altered, and Pos I inhib control, a 1,000-fold excess of the consensus p53 DNA probe without the proximity probe extension, was added to the reaction together with the full-length consensus probe. The negative control was a probe positive for HNF-4 α binding but with no known affinity for p53, and in the no lysate control, no nuclear lysate was added to the reaction. The S/N is shown on the y axis. (B) Dilutions of MCF-7 nuclear lysate were analyzed with the Pos I (filled bar) and the Pos I Mut probe (open bar) to determine the limit of detection of p53 in the extract.

p53. The tumor suppressor p53 can mediate several different functions by activating or repressing a large number of target genes. *TP53* is one of the most commonly mutated genes in human cancer, and mutations generally affect amino acids within the DNA-binding domain (21). In addition to its role in transcriptional regulation, p53 is also involved in transcription-independent regulation of apoptosis, genome integrity, DNA repair, and DNA recombination (22).

As a proof of principle, proximity ligation was used to analyze nuclear extracts from the breast carcinoma cell line MCF-7 for evidence of interactions between the DNA-binding protein p53, recognized by a p53 antibody-DNA conjugate, and partially double-stranded DNA probes including putative recognition sequences for the p53 protein. Three previously described p53-binding DNA sequences were analyzed together with specificity controls (Fig. 2A). The following sequences were shown to be positive by PLA: (i) a consensus P53-binding sequence (23), (ii) a focal adhesion kinase promoter region shown to interact with p53 (24), and (iii) a polymorphic microsatellite, (TGYCC)₁₅, that mediates induction of p53-inducible gene 3 by p53 (25). Additionally, we included the following specificity controls: (i) a modified p53 consensus sequence in which four nucleotides within the consensus binding site were altered, (ii) the above p53 consensus sequence but now analyzed in the presence of a 1,000-fold excess of consensus probe without the ligatable sequence extension, (iii) a negative control sequence shown to bind HNF-4 α , and (iv) a control for which no nuclear lysate was added to the reaction. The nuclear extract was diluted and analyzed with the consensus probe and the mutated consensus probe to determine the limit of detection, which was shown to be 2.5 ng of the MCF-7 nuclear extract (Fig. 2B). The

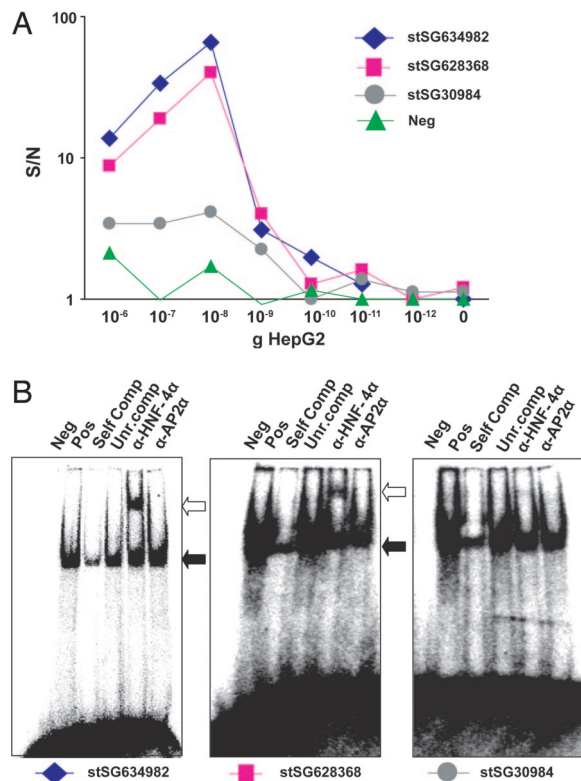


Fig. 3. PLA analysis and EMSA of tentative HNF-4 α -binding sequences. (A) Three sequences identified by ChIP-chip analysis were analyzed by PLA for interactions with HNF-4 α in HepG2 nuclear extracts together with a negative control (a p53 consensus oligonucleotide). The nuclear extract was diluted as indicated along the x axis to investigate the lowest dilution with detectable amounts of binding protein. S/N values are shown on the y axis. (B) EMSAs of sequences identified by ChIP-chip (Pos) along with negative control sequences (Neg), competitor oligonucleotides having the same (Self Comp) or an unrelated sequence (Unr.comp), and supershift reactions with HNF-4 α antibodies (α -HNF-4 α) and with irrelevant AP2 α antibodies (α -AP2 α).

experiments were additionally performed with recombinant P53 protein (Active Motif, Rixensart, Belgium) instead of the MCF-7 nuclear extract (results not shown).

HNF-4 α . Hepatocyte differentiation and metabolism are controlled both by ubiquitous and liver-specific transcription factors. HNF-4 α belongs to the nuclear receptor family and is considered the major regulator of the hepatocyte phenotype (26). Mutations in HNF-4 α resulting in an early stop codon or amino acid substitutions have been associated with an autosomal dominant form of diabetes, MODY1 (27), and polymorphisms in the HNF-4 α promoter have been linked to the common form of type 2 diabetes (28).

Three sequences identified by ChIP-chip analysis in HepG2 cells (29) and containing tentative HNF-4 α binding sites predicted by the BioProspector software (30) were analyzed by PLA in HepG2 nuclear extracts, along with a probe containing the consensus sequence for p53-binding, serving as a negative control. The nuclear extract was diluted to find the amount that resulted in the highest signal-to-noise ratio (S/N) and to identify the highest detectable dilution (Fig. 3A). The optimal total protein mass in the nuclear extract for a maximal S/N was 20 ng, and the limit of detection was \approx 1 ng, corresponding to extracts from 1 to 10 cells. Larger amounts of protein resulted in a decrease of signal, most probably because of a “high-dose hook effect,” known to occur when the amount of analyte exceeds the amount of affinity probes in an assay (31). By analyzing each consensus sequence with two or three different lysate dilutions, accurate measurements were obtained.

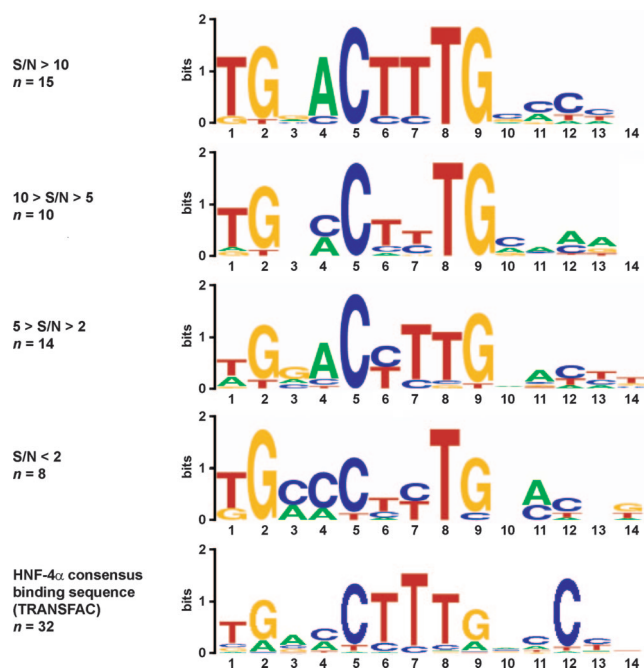


Fig. 4. HNF-4 α consensus motifs as identified by PLA and TRANSFAC. The 50 sequences shown in Fig. 4 were divided in groups based on their S/N values. The values defining each group were selected so that similar numbers of sequences were assigned to each of them. Subsequently, each of the group of sequences was analyzed by using the motif-finding program BioProspector (30), which identifies the most common motifs found among such sequences. The consensus motif found for each of the groups is illustrated with the program WebLogo (33). The consensus for each of the sequence groups represents the overall results obtained when BioProspector was run several times for each of the group, and very similar results were obtained in all cases. The previously known binding consensus sequence for HNF-4 α as found in TRANSFAC (32) is indicated at the bottom.

The PLA measurements were verified by supershift EMSA analysis of the three HNF-4 α -positive sequences (Fig. 3B), showing good agreement between the two methods. The signal generated by the DNA fragment stSG30984 was barely positive in EMSA (with \approx 4 μ g of nuclear lysate per lane), but a signal significantly higher than background was observed with the PLA, demonstrating that this assay is more sensitive than EMSA and has a wider dynamic range.

An additional 50 tentative HNF-4 α binding sequences were randomly selected from the same ChIP-chip study and analyzed by PLA in HepG2 nuclear extracts. The S/N for the proximity ligation analysis of each of the probes was compared with the previously calculated prediction score that indicates the estimated likelihood that HNF-4 α would bind to the actual sequence motifs in each of the probes [supporting information (SI) Fig. 6]. The correlation was good in that probes with the highest prediction score gave rise to high S/N, whereas probes with lower prediction score, from 4 to 7, with few exceptions demonstrated specific binding to the protein but at lower S/N.

The different probe sequences were divided in groups based on their S/N values. The S/N defining the respective groups were selected so that a similar and sufficiently high number of sequences was assigned to each of them. Subsequently, each of the groups of sequences was analyzed with the motif-finding program BioProspector (30). The consensus motif found for each of the groups together with the previously known binding consensus motif for HNF-4 α from the TRANSFAC database (32) are presented (Fig. 4) as a consensus logo created with WebLogo (33). The sequences that gave rise to the highest S/N were also those most similar to the

stranded DNA microarrays. After removing nonbound proteins by washes, the array is labeled with a fluorophore-conjugated antibody specific for the epitope tag on the expressed DNA-binding protein. The method offers high-throughput characterization of DNA–protein binding with a very high resolution, but tagged proteins may exhibit altered binding properties compared with the endogenous proteins, and exact quantification of the interactions is difficult.

In vitro selection has been successfully used to generate highly accurate descriptions of protein binding sites (42). However, SELEX will, by design of the assay, only yield “optimal” binding sequences, which might be different from those found *in vivo*. In that regard, PLA offers the advantage of allowing comparison of protein affinities for any known or natural binding sites.

The PLA technique has important advantages over existing methods for *in vitro* analysis of protein–DNA interactions. This assay is nonradioactive and has a very low reagent and sample consumption rate and extremely high sensitivity through the use of nucleic acid reporters that enable analysis of a few nanograms of nuclear extracts, corresponding to between 1 and 10 cells. Accordingly, the method is several orders of magnitude more sensitive than techniques like EMSA for investigating DNA binding of proteins, and it yields quantitative results, allowing affinities to be calculated. The assay procedure takes less than 4 h, including the PCR step. The PLA assay can be performed in a homogenous format as shown herein or on a solid support with the protein-binding DNA probe immobilized to a surface prior to interaction analysis, allowing nonspecifically bound reagents to be removed by washes (data not shown). The homogeneous assay format requires no washes or phase separations and the hands-on time is very limited, with only three additions of reagents, which makes it suitable for automation.

The required instrumentation, including the real-time PCR instrument, is standard laboratory equipment. Reagents for these assays are inexpensive and/or easy to prepare. The technique should also be suitable for multiplexing, i.e., allowing the simultaneous analysis of many different proteins and/or different DNA probe molecules. For that purpose, unique tag sequences are included in the DNA or antibody probes (Fig. 1), permitting individual amplification products to be identified on DNA tag arrays (20). Libraries of potential binding sites can be identified either by using bioinformatic approaches or by mining results from high-throughput techniques, as shown in this manuscript. As prices for oligonucleotide synthesis continue to fall, costs for synthesis of such libraries will be comparable to those charged for commercial microarray probe collections.

Proximity ligation of transiently interacting molecules may permit the analysis of interactions with affinities lower than those detectable by physical capture, and the mechanism has furthermore been shown to allow detection of complexes involving more than two components (43). The proximity ligation mechanism can also be adapted to allow *in situ* analysis of DNA–protein interactions for localized detection (O. Söderberg, personal communication), rendering the technique promising for studies of tissue heterogeneity in terms of promoter occupancy.

Cantor and coworkers (55) have developed an alternative approach for *in vitro* analysis of protein–DNA interactions. Double-stranded oligonucleotides bound by transcription factors were physically isolated and amplified, and the identities and amounts of bound sequences were identified by mass spectrometric quantification of DNA mass tags. The authors report successful multiplexing, which is limited only by the resolution defined by the length of the mass tags. We conclude that DNA molecules bound by specific proteins can be captured either by physical isolation or proximity-dependent ligation, and amplified representations of captured sequences can be identified via real-time PCR or mass spectrophotometric tags, and read-out on tag arrays could allow even greater numbers of sequences to be evaluated.

In summary, the PLA-based technique described here is suitable for highly sensitive, quantitative analyses of interactions between transcription factors and regulatory sequence elements with single-base-pair resolution. It should thus be useful for characterization of transcription factor levels across different tissues and cell types in various normal and disease states.

Materials and Methods

Antibody Biotinylation and Conjugation. Biotinylated affinity-purified polyclonal antibody against p53 was purchased from R&D Systems (Abingdon, U.K.). The HNF-4 α antibodies were purchased from Active Motif and from Santa Cruz Biotechnology (Santa Cruz, CA) (catalog no. C19). The polyclonal antibody against USF1 (H86) was purchased from Santa Cruz Biotechnology.

HNF-4 α and USF1 antibodies were biotinylated with D-biotin-N-hydroxysuccinimide (Nordic Biosite, Taby, Sweden) according to the recommendations of the manufacturer.

Streptavidin–oligonucleotide conjugates were prepared by coupling a 5' free thiol-modified oligonucleotide to maleimide-derivatized streptavidin, as previously described (19). The sequence of the 5' STV is 5'-P-TCGTGTCTAAAGTCCGTTACCTTGAT-TCCC CTAACCCTCT TGAAAAATTCGGCATCGGTGA-3'. The biotinylated antibodies were coupled with the STV-oligonucleotide conjugates as follows. The biotinylated antibodies were diluted in PBS with 1% (wt/vol) BSA (Sigma, St Louis, MO) to a final concentration of 30 nM. The antibodies were then combined with the streptavidin–oligonucleotide conjugations in a 1:1 ratio in a volume of 5 μ l at room temperature for 1 h. Thereafter, the antibody-oligonucleotide probes were further diluted to a concentration of 1.2 nM in a probe-dilution buffer [1 \times PBS/1% (wt/vol) BSA/16 μ g/ml sheared polyA bulk nucleic acid (Sigma–Aldrich, Stockholm, Sweden)/1 mM D-biotin (Molecular Probes, Eugene, OR)] and stored at 4°C until use.

DNA Probes. All DNA probes for this study were designed by using the software framework ProbeMaker (44). HPLC-purified oligonucleotide probes were purchased from Biomernet (Ulm, Germany). The common sequence, i.e., primer 2, the sequence tag, and the real-time PCR tag, as well as 12 nucleotides complementary to one half of the connector oligonucleotide (Fig. 1), is CATCGC-CCTTGGACTACGACTGACGAACCGCTTTGCCTGAC-TGATCGCT AAATCGTG.

The variable sequences contain the predicted binding sites (10–12 bp) that have been extended by 5 bp on both 5' and 3' ends.

The p53 variable sequences were as follows: p53 consensus oligonucleotide (23), TAGAGAACATGTCTAAGCATGCT-GGGGACT; p53 consensus oligonucleotide (mutated binding site), TACAGAATCGCTCTAAGCATGCTGGGGACT; “FAK-promoter region”, AGGCACCTCCGCAAGCCCCACGCAGC; and polymorphic microsatellite, (TGYCC)₁₅ (where Y is C/T).

The HNF-4 α variable sequences were as follows: stSG634982, AACTCAGGGCAAAGTTCAGCTGCTG; stSG628368, CCTCTGTAACCTTGGATTCTCAC; and stSG30984, CT-GGAAGCCAAAGACCCTAGTCAA.

The 50 additional HNF-4 α binding sequences are shown in SI Table 1.

The USF1 variable sequences were as follows: positive control (stSG627950), CCTGCCCCGTCGACCCGGCCCT; sequence 1 (stSG627950), TGCACGGTACGTCGCTCGAGC; sequence 2 (stSG609339), TGGATAGTCACTTAATGCTTA; sequence 3 (stSG628088), CATACGGTTGCGTGGATGCTC; and negative control (HNF-4 α consensus), AACTCAGGGCAAAGTTCAGCTGCTG.

The probes were made partially double-stranded by hybridizing an equimolar amount of a 20-mer oligonucleotide complementary to the variable part of the probe in 0.3 \times standard saline citrate (1 \times SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7).

Nuclear Lysates and EMSA. Nuclear lysate from peroxide-treated MCF-7 cells was purchased from Active Motif. HepG2 nuclear extracts were prepared as previously described (45). EMSAs were performed as described by Rada-Iglesias *et al.* (29).

Proximity Ligation Reaction. Incubation of samples with proximity probes p53 and HNF-4 α /USF1. For p53, DNA probes (25 pM) were incubated with 1 μ l of the nuclear extract in 10 mM Tris buffer (pH 7.5) with 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 5% (vol/vol) glycerol, and 1 μ g of poly(dI-dC) in optical PCR tubes (Applied Biosystems, Foster City, CA) in a total volume of 3 μ l for 30 min at room temperature. Two microliters of 5' free anti-p53-DNA conjugate was added to the mixture at a concentration of 20 pM, and the incubation was continued for 2 h at room temperature. The reaction conditions were adapted from a protocol provided by Santa Cruz Biotechnology.

For HNF-4 α /USF1, DNA probes (25 pM) were incubated with 1 μ l of the nuclear extract in 10 mM Hepes-KOH buffer (pH 9.7), 10% (vol/vol) glycerol, 50 mM KCl, 5 mM MgCl₂, and 0.6 mM DTT in optical PCR tubes (Applied Biosystems) in a total volume of 3 μ l for 60 min on ice. Two microliters of the 5' free anti-HNF-4 α /USF1 conjugate was added to the mixture in a concentration of 20 pM, and the incubation was continued for at least 4 h at 4°C. Reaction conditions were adapted from the EMSA protocol published by Rada-Iglesias *et al.* (29).

Ligation and quantitative real-time PCR. After the initial incubations, 45 μ l of a combined mixture for ligation and amplification was added to the 5- μ l incubations [final concentrations: 50 mM KCl,

20 mM Tris-HCl (pH 8.4), 3.15 mM MgCl₂, 0.4 Weiss units of T4 DNA ligase (Fermentas, St. Leon-Rot, Germany), 400 nM connector oligonucleotide (TACTTAGACACGACACGATT-TAGTTT; biomers.net), 80 μ M ATP, 200 μ M dNTP mixed with dUTP, 100 nM primers (forward, CATCGCCCTGGAC-TACGA; reverse, GCGAATCAAGGTAACGGACTTTAG; biomers.net), 100 nM TaqMan probe (FAM-TGACGAAC-CGCTTTGCCTGA-MGBNFQ; Applied Biosystems), and 1.5 units of Platinum TaqDNA polymerase (Invitrogen)]. After the addition of the combined ligation and amplification mix, the tubes were sealed with optical PCR lids (Applied Biosystems) and transferred to a real-time PCR instrument (Mx3000P quantitative PCR system; Stratagene, Amsterdam, The Netherlands) or a PRISM 7000 sequence detection system (Applied Biosystems). Thermal cycling conditions comprised an initial activation step of 2 min at 95°C, then 45 cycles of 15-s denaturation at 95°C, and 60 s annealing/extension at 60°C.

Analysis of the results. The samples were analyzed in duplicates or triplicates. The results presented in the figures are mean values of the S/N, where the number of ligations of proximity probe pairs in the sample are divided by the number of ligations in a negative control.

This work was supported by grants from the Swedish Research Councils for Medicine and Natural and Engineering Sciences, the Wallenberg Consortium North for Functional Genomics, the Graduate Research School in Genomics and Bioinformatics, Uppsala Bio-X, and the EU-FP6 integrated project MolTools. J.S. was supported by a Marie Curie Action Grant.

1. Ptashne M, Gann A (1997) *Nature* 386:569–577.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al.* (2001) *Science* 291:1304–1351.
3. Orlando V, Paro R (1993) *Cell* 75:1187–1198.
4. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, *et al.* (2000) *Science* 290:2306–2309.
5. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO (2001) *Nature* 409:533–538.
6. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, *et al.* (2002) *Science* 298:799–804.
7. Ng HH, Robert F, Young RA, Struhl K (2002) *Genes Dev* 16:806–819.
8. Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM (2001) *Science* 294:2357–2360.
9. Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ (2002) *Genes Dev* 16:235–244.
10. Li Z, Van Calcar S, Qu C, Cavenee WK, Zhang MQ, Ren B (2003) *Proc Natl Acad Sci USA* 100:8164–8169.
11. Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, Luscombe NM, Rinn JL, Nelson FK, Miller P, Gerstein M, *et al.* (2003) *Proc Natl Acad Sci USA* 100:12247–12252.
12. ENCODE Project Consortium (2004) *Science* 306:636–640.
13. Bulyk ML, Gentalen E, Lockhart DJ, Church GM (1999) *Nat Biotechnol* 17:573–577.
14. van Steensel B, Delrow J, Henikoff S (2001) *Nat Genet* 27:304–308.
15. Liu X, Noll DM, Lieb JD, Clarke ND (2005) *Genome Res* 15:421–427.
16. Fried M, Crothers DM (1981) *Nucleic Acids Res* 9:6505–6525.
17. Kristie TM, Roizman B (1986) *Proc Natl Acad Sci USA* 83:3218–3222.
18. Fredriksson S, Gullberg M, Jarvius J, Olsson C, Pietras K, Gustafsdottir SM, Ostman A, Landegren U (2002) *Nat Biotechnol* 20:473–477.
19. Gullberg M, Gustafsdottir SM, Schallmeiner E, Jarvius J, Bjarnegard M, Betsholtz C, Landegren U, Fredriksson S (2004) *Proc Natl Acad Sci USA* 101:8420–8424.
20. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW (1996) *Nat Genet* 14:450–456.
21. Hollstein M, Sidransky D, Vogelstein B, Harris CC (1991) *Science* 253:49–53.
22. Sengupta S, Harris CC (2005) *Nat Rev Mol Cell Biol* 6:44–55.
23. Kastan MB, Zhan Q, el-Deiry WS, Carrier F, Jacks T, Walsh WV, Plunkett BS, Vogelstein B, Fornace AJ, Jr (1992) *Cell* 71:587–597.
24. Golubovskaya V, Kaur A, Cance W (2004) *Biochim Biophys Acta* 1678:111–125.
25. Contente A, Dittmer A, Koch MC, Roth J, Dobbstein M (2002) *Nat Genet* 30:315–320.
26. Parviz F, Matullo C, Garrison WD, Savatski L, Adamson JW, Ning G, Kaestner KH, Rossi JM, Zaret KS, Duncan SA (2003) *Nat Genet* 34:292–296.
27. Yamagata K, Furuta H, Oda N, Kaisaki PJ, Menzel S, Cox NJ, Fajans SS, Signorini S, Stoffel M, Bell GI (1996) *Nature* 384:458–460.
28. Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD, Jackson AU, Deloukas P, Hunt S, Stavrides G, *et al.* (2004) *Diabetes* 53:1141–1149.
29. Rada-Iglesias A, Wallerman O, Koch C, Ameur A, Enroth S, Clelland G, Wester K, Wilcox S, Dovey OM, Ellis PD, *et al.* (2005) *Hum Mol Genet* 14:3435–3447.
30. Liu X, Brutlag DL, Liu JS (2001) *Pac Symp Biocomput* 2001:127–138.
31. Wolf BA, Garrett NC, Nahm MH (1989) *N Engl J Med* 320:1755–1756.
32. Wingender E (1988) *Nucleic Acids Res* 16:1879–1902.
33. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) *Genome Res* 14:1188–1190.
34. Pajukanta P, Lilja HE, Sinsheimer JS, Cantor RM, Lusk AJ, Gentile M, Duan XJ, Soro-Paavonen A, Naukkarinen J, Saarela J, *et al.* (2004) *Nat Genet* 36:371–376.
35. Bowen B, Steinberg J, Laemmli UK, Weintraub H (1980) *Nucleic Acids Res* 8:1–20.
36. Woodbury CP, Jr, von Hippel PH (1983) *Biochemistry* 22:4730–4737.
37. Hanes SD, Brent R (1991) *Science* 251:426–430.
38. Takahashi M, Sakumi K, Sekiguchi M (1990) *Biochemistry* 29:3431–3436.
39. Heyduk T, Lee JC (1990) *Proc Natl Acad Sci USA* 87:1744–1748.
40. Koch SJ, Shundrovsky A, Jantzen BC, Wang MD (2002) *Biophys J* 83:1098–1105.
41. Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, Snyder M, Young RA, Bulyk ML (2004) *Nat Genet* 36:1331–1339.
42. Roulet E, Busso S, Camargo AA, Simpson AJ, Mermod N, Bucher P (2002) *Nat Biotechnol* 20:831–835.
43. Soderberg O, Gullberg M, Jarvius M, Ridderstrale K, Leuchowius KJ, Jarvius J, Wester K, Hydbring P, Bahram F, Larsson LG, *et al.* (2006) *Nat Methods* 3:995–1000.
44. Stenberg J, Nilsson M, Landegren U (2005) *BMC Bioinformatics* 6:229.
45. Andrews NC, Faller DV (1991) *Nucleic Acids Res* 19:2499.
46. Vergnes L, Taniguchi T, Omori K, Zakin MM, Ochoa A (1997) *Biochim Biophys Acta* 1348:299–310.
47. Carriere V, Vidal R, Lazou K, Lacasa M, Delers F, Ribeiro A, Rousset M, Chambaz J, Lacorte JM (2005) *J Biol Chem* 280:5406–5413.
48. Kistaki E, Lacorte JM, Katrakili N, Zannis VI, Talianidis I (1994) *Nucleic Acids Res* 22:4689–4696.
49. Hung HL, High KA (1996) *J Biol Chem* 271:2323–2331.
50. Tzamelis I, Zannis VI (1996) *J Biol Chem* 271:8402–8415.
51. Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, Shahab A, Yong HC, Fu Y, Weng Z, *et al.* (2006) *Cell* 124:207–219.
52. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, *et al.* (2006) *Nat Genet* 38:431–440.
53. Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH (2004) *Cell* 119:1041–1054.
54. Kim J, Bhinge AA, Morgan XC, Iyer VR (2005) *Nat Methods* 2:47–53.
55. Zhang L, Kasif S, Cantor CR (2007) *Proc Natl Acad Sci USA* 104:3061–3066.