Identification of Estrogen-responsive Genes Using a Genome-wide Analysis of Promoter Elements for Transcription Factor Binding Sites

Received for publication, August 11, 2004, and in revised form, February 28, 2005
Published, JBC Papers in Press, March 24, 2005, DOI 10.1074/jbc.M409176200

Sitharthan Kamalakaran‡§§, Senthil K. Radhakrishnan¶¶, and William T. Beck****

*From the ‡Departments of Biochemistry and Molecular Genetics, ¶¶Medicine, Microbiology and Immunology, and ****Biopharmaceutical Sciences and Cancer Center, University of Illinois at Chicago, Chicago, Illinois 60612

We developed a pipeline to identify novel genes regulated by the steroid hormone-dependent transcription factor, estrogen receptor, through a systematic analysis of upstream regions of all human and mouse genes. We built a data base of putative promoter regions for 23,077 human and 19,984 mouse transcripts from the National Center for Biotechnology Information annotation and 8793 human and 6785 mouse promoters from the Data Base of Transcriptional Start Sites. We used this data base of putative promoters to identify potential targets of estrogen receptor by identifying estrogen response elements (EREs) in their promoters. Our program correctly identified EREs in genes known to be regulated by estrogen in addition to several new genes whose putative promoters contained EREs. We validated six genes (KIAA1243, NRP1, MADDH9, NME3, TPDP52L, and ABCG2) to be estrogen-responsive in MCF7 cells using reverse transcription PCR. To allow for extensibility of our program in identifying targets of other transcription factors, we have built a Web interface to access our data base and programs. Our Web-based program for Promoter Analysis of Genome, PAGen@UIC, allows a user to identify putative target genes for vertebrate transcription factors through the analysis of their upstream sequences. This interface allows the user to search the human and mouse promoter data bases for potential target genes containing one or more listed transcription factor binding sites (TFBSs) in their upstream elements, using either regular expression-based consensus or position weight matrices. The data base can also be searched for promoters harboring user-defined TFBSs given as a consensus or a position weight matrix. Furthermore, the user can retrieve putative promoter sequences for any given gene together with identified TFBSs located on its promoter. Orthologous promoters are also analyzed to determine conserved elements.

Estrogens play a critical role in vertebrate reproduction, especially in the development of female sex organs (1). The bulk of estrogen signaling is controlled by estrogen receptors α and β (ERα and ERβ), members of the nuclear receptor superfamily (2) of ligand-inducible transcription factors. ERα and ERβ bind to the estrogen response element (ERE) on the promoters of the genes they regulate. The ERE is a palindrome of RGTCGGA motifs separated by 3 bp (3). These elements are bound by ER dimers, with one receptor binding each motif. The sequence and spatial organization of the ERE is important for the specificity of binding. Although a crystal structure of the DNA-bound ER is available (4), the unexpected diversity of DNA response elements mediating transcriptional regulation by ERs has made finding EREs in the genome very difficult. In fact, none of the typical estrogen-responsive genes have consensus EREs.

The sequencing of the human genome (5, 6) has opened a new way to study gene regulation through bioinformatic analysis of transcription factor binding sites (TFBSs) on gene promoters. There now exist on the Web both free and subscription-based resources for such analysis of specific promoter regions to identify putative TFBSs (7–12). Given a user-defined DNA sequence, these programs can predict the potential TFBSs. These resources are valuable for researchers studying regulation of individual genes. When it comes to studying global changes in gene expression following various treatments (drugs, hormones, heat shock) of cells or animals, DNA microarrays are the most popular tool. Most of these changes in gene expression are mediated through the activation of one or more transcriptional regulators (for example, p53 in DNA damage response and steroid hormone receptors like ER and retinoic acid receptor following estrogen and retinoic acid treatment, respectively). It is possible, using bioinformatics, to predict genes that are regulated by these transcriptional regulators by analyzing the promoter regions for binding sites to these transcription factors. In one such study, Hoh et al. (10) developed a computer algorithm, p53MH, to perform a genome-wide scan for p53 binding sites and identified 2,583 genes with putative binding sites for p53. Others have used similar approaches to identify p53 target genes in promoter regions and introns and c-Myc target genes in 5′-untranslated regions, respectively (11, 13). Although these studies have shown that these approaches are worthwhile and identify potential regulated genes of transcrip-
tion factors of interest, they are limited in scope because they were applied to a single transcription factor. There is thus a need for a public resource for investigators to identify genes that are potentially regulated by specific transcription factor(s) through genome-wide bioinformatic analysis of promoter regions.

In this study we developed a pipeline to identify genes controlled by the steroid hormone transcription factor, estrogen receptor. We have also shown that our method is easily extensible to the study of all transcription factors with a characterized binding site. Our Web-based software, PAGen@UIC, described herein, allows a user to identify from a set of all annotated human and mouse promoters the ones that may be regulated by the transcription factor of interest.

EXPERIMENTAL PROCEDURES

Programs—All programming languages and software used were Open Source, supplied under a general public license. The programs were written in Python (www.python.org). We used the MySQL database server software (www.mysql.com). The application runs on an Apache 2.0 HTML server (www.apache.org). All graphs were generated using the Scipy package for Python. Programs from the EMBOSS software suite (14) were used for some functions, cpgplot for generating CpG island maps and profit for weight matrix searches.

Generation of Putative Promoter Regions and Control Random Regions—Python programs were written to generate putative promoter regions (PPRs), a sequence 2000 nt upstream and 250 nt downstream of transcription start site (TSS) in the GenBank™ annotation of the genome data (NCBI Genome Build 34 Ver1, September, 2003 for human; NCBI Genome Build 32 Ver1, March, 2004 for mouse), and written into a MySQL relational database. For control random sequences, a similar program was written to obtain 2250-nt sequences from a random position in the genome. The number of random regions per chromosome matched the number of genes in each chromosome. Orthologous information for the mouse and human genes was compiled from the ENSEMBL project (15, 16).

Cell Lines and Reagents—The MCF7 cell line was obtained from ATCC (Manassas, VA). 17-β-Estradiol was purchased from Calbiochem. Unless indicated otherwise, all other reagents and supplies were purchased from standard commercial sources.

RT-PCR—MCF7 cells were grown in phenol red-free medium supplemented with 5% charcoal dextran-stripped serum (low estrogen) for 72 h and then treated with same medium supplemented with 0, 1.0, or 10 nM 17-β-Estradiol (E2) for another 72 h. Total RNA was isolated by the TRIzol method as recommended by the manufacturer (Invitrogen); 3 μg were used for cDNA synthesis using the Thermoscript RT-PCR system (Invitrogen) and oligo(dT) primers. Custom synthesized primers were purchased from Sigma Genosys (The Woodlands, TX) and used to amplify the cDNA. PCR cycles were optimized for each gene for log phase amplification. Experiments were repeated independently at least three times.

RESULTS

The annotation of the genome by the NCBI maps the TSS of each gene in the genome (15, 17). Using this annotation, we have created a data base of sequences containing PPRs spanning 2000 nt upstream and 250 nt downstream of the TSS of each gene. The data base comprises PPRs for 23,077 human genes and 19,984 mouse genes. In addition, we obtained 8793 human and 6785 mouse promoter sequences from the Data Base of Transcriptional Start Sites (DBTSS) resource. We used the TRANSFAC data base (18, 19) of DNA binding profiles of eukaryotic transcription factors to build a regular expression (RegEx)-based consensus binding set and a position weight matrix profile for various vertebrate transcription factors. These were then run against the database of human and mouse promoters. We thus established a “PAGen data base” whereby each annotated gene is defined by a set of putative TFBSs located in its PPR. A flowchart of our approach is shown in Fig. 1.

The accuracy of prediction of a given TFBS is dependent on having a good model of transcription factor binding profile. Two major approaches, consensus set and PWMs, have been used to generate a transcription factor binding profile (see Ref. 20 for a review). The consensus method has the major disadvantage of being too rigorous (many false negatives) or too degenerate (many false positives). To overcome this problem, we have adopted a regular expression-based consensus approach (Fig. 2A). Regular expression search is an established standard for text and pattern matching (21), an important advantage being that it allows for all Boolean operations to be performed. This allows for more complex pattern searches than the use of a degenerate consensus set. It also allows for easy adoption for users familiar with some bioinformatics programming and ease of learning for others with less experience. A user can not only search for promoters with binding sites for one or more specific transcription factors but can also exclude any TFBS that the promoter set should not have.

The second approach we have taken to build the transcription factor binding profile involves generation of PWMs (Fig. 2B). The elements of a PWM correspond to scores reflecting the probability of a given nucleotide occurring at a particular position of the TFBS (see Ref. 22 for a review). PWMs provide the best approximation for determining TFBSs and have been used successfully to identify many biologically relevant targets of transcription factors (8, 10, 23).
Validation of the Promoter Set—Another factor that determines the relevance of predicted genes with the queried TFBSs is the accuracy of the TSS, because our PPRs are generated based on this information. We needed to validate the set of putative promoters to determine its usefulness, and to this end we used several tests. On a macro level, we plotted the GC content of the promoter set against a set of 2250-nt fragments randomly generated in the genome with matching gene frequencies for each chromosome. Because promoters on average are known to be GC-rich, an increase in GC content for the promoter set over a randomly generated control set demonstrated a striking bias toward the TSS in the promoter sets are included in supplemental Table I. We also plotted the number of TFBSs for a PWM-based search; (b) search the data bases for a user-defined sequence given in the form of a regular expression for a RegEx-based consensus search or a FASTA-formatted list of TFBSs for a PWM-based search; (c) retrieve putative promoter sequences for any given gene together with identified TFBSs located on its promoter. The promoter of the orthologous gene, along with CpG plots indicating putative CpG islands in sequence tag sequences to the genome and to build a dataset of proximal promoters (PromoSer; biowulf.bu.edu/lob/PromoSer/) for human, mouse, and rat genes. Pairwise sequence analysis of our set of PPRs with the corresponding regions obtained from PromoSer revealed that >90% of our promoter set had >95% identity. The DBTSS project also aims to map the exact genomic position and TSS of all human and mouse genes (12, 25). The current version, however, has only TSS for 8793 human and 6785 mouse genes. A comparison of the promoter sequences that we derived from NCBI annotation with this limited set of promoters provided in the DBTSS project showed that >85% of our promoter sequences share at least 90% identity with their respective DBTSS promoters. We found that of these remaining 15%, ~13% of the NCBI promoters were, on average, 4986 bases downstream of their corresponding DBTSS promoters (8% was found within 1 kb, 1% between 1 and 5 kb, and the rest over 10 kb). The remaining 2% were extremely far apart. To accommodate these discrepancies in promoters, we decided to use the DBTSS promoter set as an additional option.

We then searched our data base for putative estrogen-regulated genes by looking for the presence of EREs in their corresponding promoters. We were able to identify several known ER-regulated genes (EBAG9, c-fos, OXT, F12, TFF1, LTF, CTSD, PFDN2, TGF-a, AGT, GREB1). These genes and their EREs as published in the literature and confirmed using PAGen@UIC are shown in Table I. We also identified several novel candidate genes with EREs in their promoters. We tested some of these genes for estrogen regulation using RT-PCR. For our analysis we selected only those genes with an ERE occurring within 1000 nt upstream of the TSS. We chose 16 such genes, designed gene-specific primers, and examined them for regulation by estrogen in MCF7 cells following a 72-h treatment. Of the 16 genes assayed, 6 were not expressed in MCF7 cells and 4 were unresponsive to treatment. However, as shown in Fig. 4, six genes with an ERE in their upstream sequence regions were, in fact, regulated by estrogen in MCF7 cells. Five genes (KIAA1243, NRP1, MADH9, NME3, and TPD52L) were up-regulated upon treatment, whereas one (BCRP/ABCG2) was down-regulated. These genes, their predicted EREs, and position relative to their TSS are listed in Table II. The down-regulation of ABCG2/BCRP seems to be cell line-specific because it is, in fact, up-regulated following estrogen treatment in ER-positive T47D-A18 cells, an effect that is mediated by the ERE element identified by PAGen@UIC (26). Thus, by using PAGen@UIC, we have high accuracy for predicting novel estrogen-responsive genes.

Finally, we wanted our pipeline to be accessible to other researchers studying transcriptional regulation. The tools we developed can be easily extended for the study of any transcription factor whose binding site has been characterized. We built a wrapper for our data base and data mining tools so that they can be freely accessible through a browser interface on the World Wide Web.
the PPR, is also provided. The typical result of a query lists a set of genes with putative binding sites for the selected transcription factor together with the sequence and position of the identified binding site. Each identified gene is also linked to the NCBI database. For example, the user can look up the mRNA transcript of a gene, search PubMed for publications, and retrieve the chromosomal location of the gene using NCBI MapView and obtain the orthologous promoter. The number of potential target genes predicted by our program for some well-known transcription factors are included in Table III. The full view and obtain the orthologous promoter.

### Table I

| Gene | Estrogen response element and position |
|------|--------------------------------------|
| EBAG9 | GGTcAGGTTACcc, +6 |
| c-fos | CcGgCAGcGTGACC, -1216 |
| GXT | CcGTgACCtTGACC, -168 |
| F12 | AgggCACGcTGACC, -50 |
| TFF1 | AGGTACOGTTgCCa, -409 |
| LTF | AGGTACOGccGAcCT, -358 |
| CTSD | GGGcGcGcGcGcGACC, -140 |
| PFDN2 | AGGTACOGcGACC, -1776 |
| TGF-α | GTGtACCcTcGcGcC, -126 |
| AGT | GGGcCTcGCtGACC, -27 |
| GREB1 | GGTGCTCCCTACT, -1583 |

### FIG. 3

**A**, the occurrence of CpG islands against their relative position with respect to the transcriptional start site (TSS) in the PAGen and DBTSS promoter sets are contrasted with a random set. A vast majority of the promoters in the PAGen and DBTSS sets have CpG islands clustered near the TSS, whereas it is distributed throughout the length of the sequence for the random set. **B**, the position bias for binding sites of transcription factors CREB1, GC1, SP1, and CAAT in the PAGen set of promoters is compared with a random set of sequences. Transcription factors with GC-rich response elements (CREB1, GC1, and SP1) as well as non-GC-rich response elements (CAAT) tend to cluster near the TSS.

### FIG. 4

**Novel estrogen-responsive genes confirmed by RT-PCR in MCF7 cells.** Genes predicted to have EREs in their promoters were analyzed for regulation following induction with 1 or 10 nM 17β-estradiol (E2) for 72 h using RT-PCR. Five genes (KIAA1243, NRIP1, MADH9, NME3, and TPD52L) were up-regulated, and one gene (ABCG2) was down-regulated. One gene (GTF21R) that was not regulated by estrogen in MCF7 cells under the conditions tested and β-actin are shown as controls.
Table II

Novel estrogen-responsive genes identified
Novel estrogen-responsive genes validated by RT-PCR and their predicted EREs are shown.

| Gene       | ERE type   | Enron response element and position |
|------------|------------|-------------------------------------|
| KIAA1243   | Consensus  | CCGAGCAGCTCCGAC, −719               |
| NRP1       | Consensus  | GGGCGCCAGCTCCGAC, −719              |
| MADH9      | F12        | AGGGCGCCAGCTCCGAC, −342             |
| NME3       | F12        | GGGCGCCAGCTCCGAC, −916              |
| TPDS2L2    | F12        | AGGGCGCCAGCTCCGAC, −465             |
| ABCG2      | c-fos      | ACGGCAGCGGTGACAC, −531              |

A list of our transcription factors is presented in supplemental Table II.

**DISCUSSION**

In this study we have built a freely available online database of upstream elements that can be queried for binding sites for one or more transcription factors and then used it to identify and validate six novel estrogen-responsive genes. The size of the database is larger than any previously available resource. Importantly, our resource can be queried by any user-determined regular expression or PWM, allowing users to determine the expression that best describes their binding site. There are a few data bases of promoters available on the Web. The most accurate resource, the Eukaryotic Promoter database (27–29), contains annotated eukaryotic POL II promoters with an experimentally determined TSS. This data base contains only 1871 human promoters and 196 mouse promoters. This constitutes less than 10% of human genes and less than 1% of mouse genes. The two data bases similar in size to ours are PromoSer (24) and mPromDb (30). However, these tools are primarily to retrieve promoters of interest. The PromoSer and Eukaryotic Promoter data bases do not provide the capability to search for TFBSs. With mPromDb, it is only possible to retrieve a set of experimentally verified target genes of a transcription factor. By contrast, our tool, in addition to retrieving promoters of interest, also allows a user to identify a set of putative genes that may be targets of a given transcription factor through the identification of TFBSs in their promoters.

There are a few reports on genome-wide analysis of regulation by specific genes (8, 10, 11, 13, 31). These efforts concentrate on defining a specific matrix for the gene of interest. In addition some of these programs are proprietary and have not been released in the public domain. Our effort, PAGen@UIC, although, applied to one specific response element, can be adapted easily to any user-defined criteria. The freely available browser-based input gives any user the ability to search TFBSs readily for the transcription factor of interest. Moreover, our program provides the user the ability to tailor the search as needed. An investigator can make subtle changes in the consensus site based on the hypothesis being tested. Instead of defining a consensus for a given transcription factor, the user can also form a regular expression using only experimentally verified binding sites for that particular transcription factor and can find new genes that are potentially regulated by the transcription factor. For example, the tumor suppressor PTEN has been shown to be regulated by p53 (32). However, the promoter for PTEN does not contain the consensus binding site for p53 (33), which is defined as two copies of the 10-bp motif 5’-RRRC(A/T)(T/A)GYYY-3’ separated by a spacer of 0–13 bp. If we vary the spacer by one base, i.e., use 14 nucleotides instead of 13, PTEN would have been identified in these searches. In fact, we have identified many EREs in this manner. The promoter for the ABC half transporter, BCRP/ABCG2, has an ERE that is similar to the ERE of c-fos, but not the consensus ERE that has been widely published. Similarly, the MADH9 promoter has an ERE that is similar to the ERE of F12.

The RegEx-based consensus provides significantly better results than traditionally employed consensus approaches. In our case (ER binding sites), a degenerate consensus would yield 220 hits, and a rigorous consensus would yield only 3, whereas our RegEx method gave us 119 hits. The main strength of this method is that it produces a small set of hits and can reduce the number of spurious hits. However, this also means that many real hits may be omitted. PWMs, on the other hand, are the established method for identifying TFBSs. They generate many putative targets that allow one to identify genes that may not be found using the RegEx method. We believe these two approaches complement each other in fulfilling the potential needs of users.

Another important feature of PAGen@UIC is the integration of the human and mouse promoter data bases. This provides an extra level of accreditation that a promoter with a transcription factor binding site is biologically relevant. Thus, if the user searches for human promoters with binding sites for a given transcription factor, in addition to providing a list of all the human promoters that have a binding site, the program also looks for binding sites for the same transcription factor in the orthologous gene promoters in the mouse data base, and the result table will have an additional link wherever a binding site is present in the orthologous mouse promoter. For example, NRP1 is an estrogen-responsive gene we identified using this program, with a consensus ERE at position −721; by comparison the program also identified that the mouse ortholog of NRP1 has an ERE at position −651. This conservation of the ERE increases the probability that it is a relevant hit. This functionality in PAGen@UIC is especially relevant if the transcription factor binding site is poorly defined (e.g., c-Myc has a consensus binding site, CACGTG) and the search generates many hits. In such cases, the user can focus only on those hits where the promoters of both the mouse and human genes are predicted to have the binding site for the desired transcription factor.

Although we project that PAGen@UIC will be a useful tool, we are also aware that there are some limitations. First, our data base of upstream sequence regions is based on the NCBI annotation of the transcription start site for each gene. The tests we have performed show that the promoter set based on NCBI annotation is reliable and compares well with other promoter data bases. However, it is possible that there are a few errors in the annotation that will negatively impact our data set. Second, we have not undertaken any efforts to determine the exact regulatory region for each gene. Our data base of −2000 to +250 base sequences contains only those potential regulatory regions within this region. Although this will hold true in most cases, there are notable exceptions. For example, we know that the p53 binding site for p21 is more than 2 kb upstream of the promoter. We have chosen not to include sequences further upstream because their inclusion will add greatly to the numbers of false positives while detecting few relevant binding sites. PAGen@UIC cannot identify genes with non-canonical transcription factor binding sites or genes that are indirectly regulated by transcription factors (with no binding site in the regulatory regions). There have also been several reports of transcription factors binding introns and 5’-untranslated regions. The 5’-untranslated regions have not been included in this data base. However, a separate data base of human and mouse 5’- and 3’-untranslated regions with all the functionalities described here for such searches is currently under development.

Our tool provides an integrated approach to the study of gene regulation through the analysis of upstream promoter ele-
ments. The number of promoters in our data base compares favorably to the other available resources. A significant addition provided by our resource enables the user to identify putative targets of various transcription factors. The ability to perform both regular expression searches and weight matrix searches on a large set of putative promoters is unique to our tool. The integration of the mouse and human promoters in every search operation to identify conservation of the binding sites adds greatly to the usefulness of our tool. User-defined searches of our promoters will allow researchers to modify TFBs where the TRANSFAC profile is not deemed to be accurate.

Further, this approach can potentially provide functional characterization for previously unknown or predicted genes. For example, the KIAA1243 gene is a predicted gene with some expressed sequence tag evidence, but it has no characterized function. We were able to identify it as an estrogen-responsive gene using our application, thus providing a direct approach to characterize its function. We thus expect our application to be quite useful for the study of gene regulation and development pathways associated with one or more transcription factors.

Acknowledgments—We thank Drs. Yin Yuan Mo, Xiaolong He, Rachel Ee, and Martina Vaskova for critical comments and suggestions.

REFERENCES

1. Rosenfeld, C. S., Wagner, J. S., Roberts, R. M., and Lubahn, D. B. (2001) Science 291, 2304–2312.

2. Schwabe, J. W., and Teichmann, S. A. (2004) Genome-wide Identification of Transcription Factor Targets

3. Sanchez, R., Nguyen, D., Rocha, W., White, J. J., and Mader, S. (2002) BioEssays 24, 244–254

4. Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993) Cell 75, 567–578

5. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Haeford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McCombie, W., Mardis, E., McPherson, J., Nemoto, T., Nelson, W., Pfafferott, M., Reich, D., Rogers, Y. H., Shriver, D., Stojanov, P., and Voelkel, J. (2001) Nature Genet. 29, 25–29

6. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Gocayne, J. D., Jones, T. J., et al. (1995) Science 268, 201–211

7. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) Nucleic Acids Res. 23, 1078–1080
Genome-wide Identification of Transcription Factor Targets

1. Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohmhauser, R., Pruss, M., Schacherer, F., Thiele, S., and Urbach, S. (2001) Nucleic Acids Res. 29, 281–283

2. Bulyk, M. L. (2003) Genome Biol. 5, 201

3. Friedl, J. E. F. (2002) Mastering Regular Expressions, 2nd Ed., O'Reilly and Associates, Sebastopol, CA

4. Rice, P., Longden, I., and Bleasby, A. (2000) Trends Genet. 16, 276–277

5. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Potter, M., Poter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Staliker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., and Clamp, M. (2002) Nucleic Acids Res. 30, 38–41

6. Kasprzyk, A., Keefe, D., Smedley, D., London, D., Spooner, W., Melsopp, C., Hammond, M., Roza-Serra, P., Cox, T., and Birney, E. (2004) Genome Res. 14, 160–169

7. Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Tatusova, T. A., and Wagner, L. (2000) Nucleic Acids Res. 28, 33–33

8. Wingender, E., Dietze, P., Karas, H., and Knuppel, R. (1996) Nucleic Acids Res. 24, 359–371

9. Bajic, V. B., Tan, S. L., Chong, A., Strom, A., Gustafsson, J. A., Lin, C. Y., and Liu, E. T. (2003) Nucleic Acids Res. 31, 3605–3607

10. Kel, A. E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O. V., and Wingender, E. (2003) Nucleic Acids Res. 31, 3576–3579

11. Hoh, J., Jin, S., Parrado, T., Edington, J., Levine, A. J., and Ott, J. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 8467–8472

12. Wang, L., Wu, Q., Qiu, P., Mirza, A., McGuirk, M., Kirschmeier, P., Greene, J. R., Wang, Y., Pickett, C. B., and Liu, S. (2001) J. Biol. Chem. 276, 43604–43610

13. Suzuki, Y., Yamashita, R., Sugano, S., and Nakai, K. (2004) Nucleic Acids Res. 32, (Data base issue) D78-D81

14. Schuldiner, O., Shor, S., and Benvenisty, N. (2002) Gene 292, 91–99

15. Rice, P., Longden, I., and Bleasby, A. (2000) Trends Genet. 16, 276–277

16. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Potter, M., Poter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Staliker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., and Clamp, M. (2002) Nucleic Acids Res. 30, 38–41

17. Kasprzyk, A., Keefe, D., Smedley, D., London, D., Spooner, W., Melsopp, C., Hammond, M., Roza-Serra, P., Cox, T., and Birney, E. (2004) Genome Res. 14, 160–169

18. Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Tatusova, T. A., and Wagner, L. (2000) Nucleic Acids Res. 28, 33–33

19. Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohmhauser, R., Pruss, M., Schacherer, F., Thiele, S., and Urbach, S. (2001) Nucleic Acids Res. 29, 281–283

20. Bulyk, M. L. (2003) Genome Biol. 5, 201

21. Friedl, J. E. F. (2002) Mastering Regular Expressions, 2nd Ed., O'Reilly and Associates, Sebastopol, CA

22. Stormo, G. D. (2000) Bioinformatics 16, 16–23

23. Benes, V. P., Bulyk, M. L., and Stormo, G. D. (2002) Nucleic Acids Res. 30, 4442–4451

24. Halees, A. S., Leyfer, D., and Weng, Z. (2003) Nucleic Acids Res. 31, 3554–3559

25. Suzuki, Y., Yamashita, R., Nakai, K., and Sugano, S. (2002) Nucleic Acids Res. 30, 328–331

26. Ee, P. L., Kamalaharan, S., Tonetti, D., He, X., Ross, D. D., and Beck, W. T. (2004) Cancer Res. 64, 1247–1251

27. Cavin Perier, R., Junier, T., and Bucher, P. (1998) Nucleic Acids Res. 26, 353–357

28. Perier, R. C., Praz, V., Junier, T., Bonnard, C., and Bucher, P. (2000) Nucleic Acids Res. 28, 302–303

29. Schmid, C. D., Praz, V., Delorenzi, M., Perier, R., and Bucher, P. (2004) Nucleic Acids Res. 32, D82-D85

30. Sun, H., and Davuluri, R. V. (2004) Bioinformatics 20, 727–734

31. Bourdeau, V., Deschenes, J., Metivier, R., Nagai, Y., Nguyen, D., Bretschneider, N., Gannon, F., White, J. H., and Mader, S. (2004) Mol Endocrinol. 18, 1411–1427

32. Stambolic, V., MacPherson, D., Sas, D., Lin, Y., Snow, B., Bang, Y., Benchimol, S., and Mak, T. W. (2001) Mol. Cell 8, 317–325

33. el-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W., and Vogelstein, B. (1992) Nat. Genet. 1, 45–49