**THE ESRG-TRRD: DATABASE OF GENES WITH SPECIFIC TRANSCRIPTION REGULATION IN ERYTHROID CELLS.**

PODKOLODNAYA O.A.1, STEPANENKO I.L.

Laboratory of Theoretical Genetics, Institute of Cytology and Genetics (Siberian Branch of the Russian Academy of Sciences), Lavrentieva 10, Novosibirsk, 630090, Russia;

1Corresponding author e-mail:opodkol@bionet.nsc.ru

Keywords: database, transcription regulation, erythroid cells, differentiation, vertebrate ontogenesis, gene expression, transcription factor

The ESRG-TRRD database, compiling the information on the genes with specific transcription regulation expressed in erythroid cells, is described. The genes included in the database may be divided into certain major groups: globins, transcription factors, enzymes of the heme biosynthetic pathway, etc. All the genes contain regulatory regions responsible for erythroid-specific transcription regulation. Occurrence of alternative promoters is a typical peculiarity of many nonglobin genes described in the database. The transcription factors GATA-1, NF-E2, EKLF, PREIIBF, SSP, BGP1, b-DRF, NF-E3, and NF-E4 determine the tissue- and stage-specificity of gene transcription regulation in erythroid cells.

**Introduction**

The ESRG-TRRD (Erythroid-Specific Regulated Genes) database is a constituent of the TRRD (Transcription Regulatory Regions Database [A.E. Kel. et al., 1997; N.A. Kolchanov et al., 1998] and was designed to accumulate the information on the genes expressed in the erythroid cells in the course of their differentiation at different stages of vertebrate ontogenesis. This information forms two blocks: the data on gene expression and structure-function organization of their regulatory regions. Peculiarities of gene expression in different cell types, organs, and tissues depending on developmental stage of the organism and effect of external factors are described in the former block along with the references to the gene regulatory regions and transcription factor binding sites involved in realization of a particular expression pattern. The latter block describes the gene regulatory regions, transcription factor binding sites, and the factors binding to these sites. The database also includes a system of cross-referenced hypertexts available to the Internet users at<http://www.bionet.nsc.ru/trrd/>.

**Contents of the ESRG-TRRD database**

The ESRG-TRRD database contains description of 42 genes including, first, the genes coding for erythroid-specific proteins, such as globins, as well as the genes with a less restricted expression, the transcription of which is regulated specifically in erythroid cells. Genes encoding a number of enzymes, transcription factors, and cell surface antigens represent the latter group. In addition, the genes encoding the protein products involved in erythroid differentiation and their receptors are also included into the database. Listed in Table 1 is the information on the genes contained in the ESRG-TRRD: gene name, organism name, accession number of gene in the TRRD, number of regulatory regions and transcription factor binding sites described in the database, and EMBL accession number of the gene.

The ESRG-TRRD database contains a formalized description of 92 regulatory units (promoters, silencers, enhancers), and 411 transcription factor binding sites. Analysis of the information contained in the ESRG-TRRD database allowed us to reveal a multilevel organization of the regions involved in transcription regulation of erythroid-specific genes. The locus control region (LCR), a regulatory region common for all the genes of one globin cluster, is the highest regulatory level described in this database. This regulatory region aids the tissue-specific regulation of all the genes of the cluster and their switching in the course of ontogenesis. LCRs are marked with DNase I hypersensitive sites, which may be located in 5'- and 3'- regions of globin clusters. A minimal functional domain of beta-globin cluster LCR hypersensitive site includes GATA, AP-1/NF-E2, and CACC binding sites for erythroid-specific transcription factors GATA-1 and NF-E2 and transcription factor Sp1. Functional features of alpha- and beta-globin cluster LCRs differ something, possibly depending on chromatin structure of a - and b -globin clusters.

The next regulation level is represented by the elements providing gene transcription initiation and modulation of transcription - promoters, silencers, and enhancers. These elements were designated as regulatory units and may have tissue- and stage-specific characteristics.

Table 1. The genes described in the ESRG-TRRD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Gene name | Organism | TRRD | Regulatory units | Binding sites | EMBL AC |
| aA-globin | chicken | G000046 | 2 | 13 | X56801 |
| alpha-D-globin | chicken | G000047 | 1 | 3 | V00411 |
| pi-globin | chicken | G001169 | 1 | 7 | V00408 |
| betaA-globin | chicken | G000051 | 2 | 24 | V00409 |
| betaH-globin | chicken | G000053 | 1 | 7 | X13142 |
| rho-globin | chicken | G001184 | 4 | 8 | L17432 |
| alpha1-globin | human | G000195 | 2 | 13 | J00153 |
| alpha2-globin | human | G001202 | 2 | 8 | J00153 |
| zeta2-globin | human | G000419 | 2 | 9 | J00182 |
| beta-globin | human | G000215 | 4 | 21 | U01317 |
| delta-globin | human | G000240 | 1 | 2 | U01317 |
| Agamma globin | human | G000261 | 3 | 32 | M91037 |
| Ggamma-globin | human | G001095 | 1 | 5 | V00517 |
| epsilon-globin | human | G000253 | 3 | 8 | V00508 |
| alpha-globin | rabbit | G001076 | 5 | 12 | M74142 |
| alpha1-globin | mouse | G000468 | 2 | 15 | J00410 |
| zeta-globin | mouse | G001223 | 1 | 2 | X62302 |
| beta major -globin | mouse | G000484 | 5 | 12 | M18646 |
| porphobilinogen deaminase | human | G000359 | 2 | 10 | M18800 M95623 |
| porphobilinogen deaminase | mouse | G000584 | 4 | 12 | M28663 |
| delta-aminolevulinate dehydratase | human | G001186 | 2 | - | X64467 |
| L-type pyruvate kinase gene | rat | G000763 | 3 | 25 | X85791 |
| glutathion peroxidase | mouse | G001229 | 2 | 11 | X03920 |
| 15-lipoxygenase | rabbit | G001274 | 2 | 12 | X90860 |
| carbonic anhydrase I | mouse | G001260 | 2 | 3 | - |
| carbonic anhydrase I | human | G001261 | 2 | 8 | - |
| HOXB2 | human | G001216 | 1 | 2 | X78978 |
| GATA1 | mouse | G000512 | 2 | 3 | X56854 |
| GATA1 | chicken | G000060 | 1 | 13 | M59937 |
| TAL1 | human | G001188 | 1 | 6 | S53698 |
| TAL1 | mouse | G001189 | 2 | 7 | U01530 |
| erythroid Kruppel-like factor | mouse | G001081 | 1 | 5 | - |
| erythropoietin | human | G000254 | 2 | 5 | M11319 |
| erythropoietin | mouse | G001221 | 1 | 2 | M12930 |
| erythropoietin receptor | mouse | G000507 | 3 | 4 | M38133 |
| erythropoietin receptor | human | G001222 | 4 | 11 | M77244 |
| glycophorin A | human | G001082 | 1 | - | M24123 |
| glycophorin B | human | G000275 | 1 | 8 | M57233 |
| glycophorin C | human | G000276 | 1 | 1 | X14242 |
| glycoprotein D | human | G001262 | 1 | 2 | X85785 |
| transferrin receptor | human | G000406 | 1 | 4 | X05339 |
| histone H5 | chicken | G000063 | 4 | 19 | J00870X00169 |
| alpha locus control region | human | G001085 | 1 | 10 | S49899 |
| alpha locus control region | mouse | G001185 | 1 | 12 | U08220 |
| beta-LCR | human | G000328 | 2 | 15 | - |

A minimal promoter of globin genes providing a low level of constitutive expression of a reporter construct comprises, as a rule, САССС, ССААТ, and ТАТА boxes. Occurrence of functional GATA sites, binding with GATA-1 factor, determines the tissue-specificity of transcription from most globin promoters. Each of the globin genes has both positive and negative regulatory elements; together with LCR, these elements provide a high level of erythroid-specific transcription and gene switching in the course of ontogenesis.

Occurrence of alternative erythroid-specific promoters is a peculiarity of most genes of enzymes and transcription factors (10 of 14) contained in the ESRG-TRRD database. The distances between the alternative promoters differ considerably in the genes described. For example, erythroid- and colon-specific promoters in human carbonic anhydrase I gene (G0012601) are separated by 36 kb intron, whereas the distance between the alternative promoters in mouse TAL-1 gene (G001189) is about 300 bp. Transcription activity of the genes with erythroid-specific regulation is modulated by positive and negative regulatory elements, enhancers and silencers, providing transcription increase in erythroid cells by positive regulatory elements, as for example, in human porphobilinogen deaminase gene (G000584)[ C. Porcher, et al 1991], or its suppression by the silencers in non-erythroid cells, as in rabbit 15-lipoxygenase gene (G001274)[O'Prery and P.R. Harrison1995]. Among 46 nonpromotor regulatory units described in the database, 22 are positive; 11, negative.

Interaction of the transcription factor with the corresponding binding site is an elementary event underlying the transcription regulation. Over 400 transcription factor binding sites are described in the ESRG-TRRD database. The corresponding transcription factors are identified for 227 of the binding sites (110, broadly expressed; 117, erythroid). Among commonly expressed factors, the most of binding sites were revealed for transcription factor Sp1 (38 sites). The group of erythroid factors is represented by factors GATA-1, NF-E2, EKLF, PREIIBF, SSP, BGP1, b-DRF, NF-E3, and NF-E4. Factors EKLF, PREIIBF, SSP, and NF-E4 are involved in regulation of stage-specific globin expression.



Fig 1. The role of transcription factor GATA-1 in the terminal erythroid differentiation ( rectangle – gene; ellipse – protein; arrow – activation )

Binding of GATA-1 and NF-E2 transcription factors determines in most cases the tissue-specific gene expression. The ESRG-TRRD database describes 94 GATA-1 binding sites. Note that binding sites for this factor were recorded in virtually all genes compiled in the ESRG-TRRD. Occurrence of such sites in the genes of factor GATA-1 itself provides a possibility of positive autoregulation of these genes (Fig. 1). The second circuit of GATA-1 positive autoregulation is connected with the presence of GATA-1 binding sites in the erythropoietin receptor gene (G000507, G001222) promoter. Although the details of the EpoR-based GATA-1 positive autoregulation mechanism are vague, the process may contribute considerably to the increase in GATA-1 level. The increase in GATA-1 level, in turn, stimulates the transcription of various erythroid-specific genes (Fig. 1) due to the presence of GATA-1 binding sites in their regulatory regions [Podkolodnaya and Stepanenko, 1997]. In addition, GATA-1 stimulates transcription of the genes coding for transcription factors EKLF, TAL1, RBTN2, and HOXB2 [Crossley et al., 1994; Lecointe et al., 1994; Royer-Pokora et al., 1990; Vieille-Grosjean et al., 1995]. These factors, in turn, stimulate additionally the erythroid-specific genes through their binding sites located in the regulatory regions of the genes. These processes result in a shift towards terminal erythroid differentiation accompanied with a high expression of late erythroid-specific genes.

**Conclusion**

The ESRG-TRRD database, compiling the available information on expression and transcription regulation peculiarities of the genes expressed at different developmental stages in vertebrate erythroid cells of which transcription is regulated specifically during erythropoiesis, is described. Characteristic of the genes coding for globins is a more complex structure-function hierarchy of the regulatory regions providing the tissue- and stage-specific regulation of their transcription. Occurrence of alternative promoters is characteristic of many genes of enzymes and transcription factors which are expressed in erythroid cells. Further development of this work will involve not only the expanding of the database, but also the analysis of relations between expression patterns and structure of regulatory regions of the genes described and construction of the gene network for this functional gene group.

**Acknowledgments**

This work was partially supported by the Russian State Human Genome Program (12312 GCh-5) and Russian Foundation for Basic Research (grants 97-04-49740 and 96-04-50006).

**References**

1. Kolchanov, E.V. Ignatieva, O.V. Kel-Margoulis, A.E. Kel, E.A. Ananko., O.A. Podkolodnaya, I.L. Stepanenko, T.I. Merkulova, T.N. Goryachkovskaya, F.A. Kolpakov, N.L. Podkolodny, S.V. Lavryushev, D.A. Grigorovich, A.S. Frolov, A.G. Romashchenko, Transcription regulatory regions database (TRRD): new possibilities provided by release 4.0 this issue
2. Kel A.E., N.A. Kolchanov, O.V. Kel, A.G. Romashchenko, E.A. Anan'ko, E.V. Ignatieva, T.I. Merkulova, O.A. Podkolodnaya, I.L. Stepanenko, A.V. Kochetov, F.A. Kolpakov, N.L. Podkolodnyi, and A.N. Naumochkin, "TRRD: database on transcription regulatory regions of eukaryotic genes" Mol. Biol. (Msk.) 31, 521-530 (1997).
3. Porcher, G. Pitiot, M. Plumb, S. Lowe, H. de Verneuil and B. Grandchamp, «Characterization of hypersensitive sites, protein-binding motifs, and regulatory elements in both promoters of the mouse porphobilinogen deaminase gene» J. Biol. Chem. 266, 10562-10569 (1991).
4. O'Prery and P.R. Harrison, «Tissue-specific regulation of the rabbit 15-lipoxygenase gene in erythroid cells by a transcriptional silencer» Nucleic Acids Res. 23, 3664-3672 (1995).
5. Crossley, A.P. Tsang, J.J. Bieker., and S.H. Orkin, «Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1» J. Biol. Chem. 269, 15440-15444 (1994).
6. Lecointe, O. Bernard, K. Naert, V. Joulin, C.J. Larsen, P.H. Romej and Mathieu-Mahul D., «GATA-and SP1-binding sites are required for the full activity of the tissue-specific promoter of the tal-1 gene» Oncogene 9, 2623-2632 (1994).
7. Royer-Pokora, M. Rogers, T.H. Zhu, S. Schneider, U. Loos and U. Beolitz, «The TTG-2/RBTN2 T cell oncogene encodes two alternative transcripts from two promoters: the distal promoter is removed by most 11p13 translocations in acute T cell leukaemia's (T-ALL)» Oncogene 10, 1353-1360 (1990).
8. Vieille-Grosjean and P. Huber, «Transcription factor GATA-1 regulates human HOXB2 gene expression in erythroid cells» J. Biol. Chem. 270, 4544-4550 (1995).