Cloning, expression analysis and chromosomal localization of the human nuclear receptor gene GCNF

I.Y. Agoulnik[a], Y. Cho[b], C. Niederberger[b], D.G. Kieback[a,c,*], A.J. Cooney[c]

[a] *Department of Obstetrics and Gynecology, 8550 Fannin, Houston, TX 77005, USA*
[b] *Department of Urology, University of Illinois at Chicago, 840 South Wood St., Chicago, IL 60612, USA*
[c] *Department of Cell Biology, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA*

Received 9 January 1998

Abstract Germ cell nuclear factor (GCNF) is an orphan member of the nuclear receptor gene superfamily. We report the cloning of a cDNA encoding a new variant of human GCNF from human testis and its expression analysis. Southern blot analysis of the human genomic DNA indicates that the GCNF gene is not closely related to other members within the nuclear receptor superfamily. Chromosomal localization of the GCNF gene shows that the gene is located on chromosome 9 at the locus q33–34.1. In situ hybridization analysis of GCNF expression in the testis shows that human GCNF is expressed exclusively in germ cells.

Key words: Orphan receptor; Germ cell nuclear factor

1. Introduction

Members of the nuclear receptor gene superfamily, which are ligand-activated transcription factors, are involved in the regulation of development, growth, and reproduction [1–3]. Orphan receptors form a part of the nuclear receptor gene superfamily [3–7]. Structurally, orphan receptors share several common functional traits with the steroid receptors. Most of the orphan receptors have a DNA binding domain (DBD) and a putative ligand binding domain (LBD), although there are a few exceptions to this general rule. For example, DAX1 contains only a LBD related to nuclear receptors [8], whereas the *Drosophila* orphan receptor Knirps [9] and related genes contain only a DBD. For the majority of the orphans, ligands, and therefore possible mechanisms of activation, remain unknown. These so-called orphan nuclear receptors are found in all metazoan species, and some of them will likely interact with novel ligands, while others may be activated by post-transcriptional modifications in ligand-independent manners or by heterodimerization with active receptor partners, e.g. LXR and RXR [10].

It has been shown that orphan receptors play important physiological roles in the development, differentiation and homeostasis of many organisms. For example, FTZ-F1 null mutant mice lack gonads and adrenal glands and die during the first days of life, probably due to the absence of the adrenal function [11–14]. Loss-of-function mutations of DAX-1 have been identified as the genetic cause of the X-linked adrenal hypoplasia congenita and X-linked hypogonadotropic hypogonadism, characterized by structural abnormalities of the adrenal glands and gonads, resulting in impaired steroidogenesis [8,15]. HNF4 expression in the adult is restricted to the liver, kidney, and intestine [16]. However, HNF4 null mutants die at approximately 8.5 dpc and exhibit extensive cell death in the ectoderm at 6.5 dpc [17]. The orphan receptors COUP-TFI and COUP-TFII are transcriptional repressors that have been highly conserved during evolution [18–21]. COUP-TFI mutants die perinatally, and COUP-TFII mutants die in utero [22]. Dramatic phenotypes of the null mutants of the orphan receptors illustrate their importance in every step of development and functioning of an organism.

Germ cell nuclear factor (GCNF) is a novel orphan receptor that is not closely related to other family members [23–25]. The expression of mouse GCNF (mGCNF), in the adult, is predominantly germ cell specific [23]. It is expressed in the oocytes of growing follicles in females and after the completion of the meiosis in round spermatids in males [23,26]. High level of expression of the human homolog was detected by the Northern blot analysis only in testis and lower levels of expression were found in prostate ovary and colon [27,28]. The smaller message is 2.2 kbp in the human whereas it is 2.4 kbp in the mouse. The major difference between the size of the two messages is due to the N-terminal exon variations and the length of the 3′ untranslated region. It is known that the differences in the two messages in the mouse lies in the 3′ untranslated region [26]. The different 3′ untranslated regions may allow differential translation of the messages, which in the testis is an important mechanism for the regulation of expression [29]. Since the two messages are maintained in both species it implies that the possible post-transcriptional regulation of GCNF expression in germ cells may be important in regulating its function. The very specific pattern of expression suggests an important function for GCNF in germ cell differentiation. To investigate the pattern of expression and possible role of the GCNF in human gametogenesis we have cloned cDNA encoding the human GCNF (hGCNF) gene using the mGCNF cDNA as a probe. In situ hybridization studies show that the expression of GCNF in testis is restricted to germ cells, the spermatids and late stage spermatocytes.

2. Materials and methods

2.1. Cloning of the complementary DNA and sequencing

10^6 clones from an adult human testis random primed cDNA library (Clontech, Palo Alto, CA) were plated out and screened at low stringency with a 2 kbp cDNA fragment (nucleotides 189–2602) encoding mouse GCNF. Hybridizations were performed at 60°C overnight in Church buffer (0.5 M sodium phosphate pH 7.4, 7% SDS) and were followed by three washes at 60°C for 30 min in 1×SSC:0.1% SDS. The Marathon Ready cDNA kit (Clontech, Palo Alto, CA) was used to clone the missing 5′ end of the open

*Corresponding author. Fax: (1) (713) 798-5333.
E-mail: dkieback@bcm.tmc.edu

PII S0014-5793(98)00142-2

reading frame (ORF) of the GCNF gene, according to the manufacturer's protocol. The gene specific primer used for the amplification was PN2 CAGCTGGTGAATCAGACTGT positioned between nucleotides 879–860. The PCR fragments generated by this procedure were subcloned into pCRII vector (Invitrogen) and double-stranded dideoxy sequenced using Sequenase 2.0 kit (Amersham) and several forward and reverse primers.

2.2. Southern blot analysis

Southern blot analysis was performed using a Human GENOBLOT (Clontech, Palo Alto, CA) Southern blot of human genomic DNA digested with the *Bgl*II, *Pst*I, *Bam*HI, *Hin*dIII, and *Eco*RI restriction endonucleases. The probe used was hNF5 cDNA between 803–1860 nucleotides hGCNF. The hybridization was performed in Church buffer overnight at 65°C and was followed by three 30 min washes in 1×SSC, 0.1% SDS at 65°C, and one 30 min wash in 0.1×SSC, 0.1% SDS at 65°C. The filter was exposed to X-ray film (Kodak XAR-5) overnight.

2.3. Radioactive in situ hybridization

Testis biopsies were obtained from brain dead organ donors donating tissue for transplantation and research, in compliance with the University of Illinois at Chicago Institutional Review Board human and animal protocols, and the Regional Organ Bank of Illinois. Tissue was fixed by immersion for 18 h in the fixative that includes 3.5% paraformaldehyde, 0.15% glutaraldehyde in phosphate buffer (pH 7.2) with 2% added sucrose. After fixation, the tissue was washed briefly with phosphate buffer, dehydrated in graded ethanol series, and paraffin embedded for sectioning (3–4 μm).

The probe used was the part of the hGCNF clone spanning nucleotides 803–1860 of hGCNF. GCNF antisense and sense (negative control) riboprobes were generated from *Bam*HI and *Xho*I linearized GCNF in pCRII (Invitrogen). The Gemini System II (Promega) kit was used with T7 RNA polymerase site 1 and with the *Bam*HI linearized plasmid to generate antisense probe, and with Sp6 RNA polymerase with the *Xho*I linearized plasmid to generate sense probe. P-33 was used for all nucleotide labeling.

Sections were dewaxed in xylene and rehydrated with graded ethanol, then digested in 0.001% proteinase K at 37°C for 15 min and then acetylated with a 0.3% volume/volume solution of acetic anhydride in 0.1 M triethanolamine (pH 8.0), rinsed in 2×SSC and dehydrated in graded ethanol. Sections were then hybridized in 10^7 dpm/ml riboprobe in a buffer of 50% formamide, 10% dextran sulfate, 0.3 M NaCl, 1×Denhardt's, 10 mM Tris (pH 8.0), 1 mM EDTA (pH 8.0), 0.5 mg/ml *Escherichia coli* tRNA and 10 mM DTT for 20 h at 56°C. Slides were then rinsed in 4×SSC and digested in 0.02 mg/ml RNase A for 30 min at 37°C, gradually desalted in SSC with 1 mM DTT, and incubated in 0.1×SSC with 1 mM DTT for 30 min at 60°C. Slides were dehydrated in graded ethanol, then dipped in photographic emulsion (Kodak NTB-2), exposed for 8 weeks at 4°C in total darkness, developed, and counterstained with hematoxylin and eosin.

2.4. Chromosomal localization

Human BAC genomic DNA library was screened by PCR with GCNF specific primers PN12 (CATCTTCAATCAACATGGGCG) and PN13 (TCTGTGGAACTGAATGGATTCATG). The resulting BAC clone was used for preparation of the non-isotopically labeled probe and fluorescent in situ hybridization [30].

3. Results

3.1. Cloning of the human GCNF cDNA

A human testis complementary DNA library (Clontech, Palo Alto, CA), made from the testis of four adult healthy males, was screened with a ^{32}P-labeled probe containing the mGCNF cDNA sequence. Identified clones were amplified with the primers specific to the vector in which the library was constructed (λgt11), then subcloned into the pCRII vector (Invitrogen, Carlsbad, CA), and subsequently sequenced. The sequences were analyzed using the Bestfit program from the GCG package (Wisconsin Package Version 9.0-UNIX). Sequence analysis of the five clones obtained from the screen of the human testis cDNA library showed that they were overlapping clones with identical sequence in the regions of overlap. Sequence comparison showed that these clones were highly homologous to the mGCNF, however all of them lacked the 5′ end of the ORF and the 5′ untranslated region. The longest clone, hNF5, spanned nucleotides 803–1860 of hGCNF. To obtain a full length copy of the hGCNF cDNA a Marathon Ready cDNA library was amplified with a hGCNF specific primer PN2 and the adapter primer AP2 supplied in the Marathon Ready cDNA kit (PT1156-1, Clontech, Palo Alto, CA). The hGCNF cDNA ORF encodes a 454 amino acid polypeptide with a predicted molecular weight of 52 kDa.

At least three different variants of hGCNF have been cloned in different laboratories (Fig. 1) [27,31]. All variants are identical in the DBD and LBD and thus do not represent orphan receptor isoforms encoded by separate genes, but rather represent differential intron/exon splicing of pre mRNA derived from a single hGCNF gene. The hGCNF cDNA we have cloned has a 22 amino acid deletion in the N-terminus of the domain relative to the other three variants. It also has a four amino acid deletion found in one of the other variants (S83309). The functional significance of these N-terminal variations is, at present, unknown, but may lead to differential transactivation or DNA binding properties.

3.2. Analysis of the hGCNF gene structure and chromosomal localization

Southern blot analysis of the human genomic DNA showed that the hGCNF probe hybridizes as a single band with the

```
          1                                                    50
AF004291  MERDE~~~~~ ~~~~~~~~~~ ~~~~~~~PPP PRNGFCQDEL AELDPGT...
s83309    MERDEPPPSG GGGGGGSAGF LEPPAALPPP PRNGFCQDEL AELDPGT...
hsu64876  MERDEPPPSG GGGGGGSAGF LEPPAALPPP PRNGFCQDEL AELDPGTISV
hsu80802  MERDEPPPSG GGGGGGSAGF LEPPAALPPP PRNGFCQDEL AELDPGTISV

          51                                                  100
AF004291  .nDRAEQRTC LICGDRATGL HYGIISCEGC KGFFKRSICN KRVYRCSRDK
s83309    .nDRAEQRTC LICGDRATGL HYGIISCEGC KGFFKRSICN KRVYRCSRDK
hsu64876  SdDRAEQRTC LICGDRATGL HYGIISCEGC KGFFKRSICk KRVYRCSRDK
hsu8080   SdDRAEQRTC LICGDRATGL HYGIISCEGC KGFFKRSICN KRVYRCSRDK
```

Fig. 1. Comparative analysis of the amino acid sequence of hGCNF. Comparison of the first 100 amino acids of hGCNF variants. The first 100 amino acids contain the amino-terminal domain and part of the DNA binding domain. Four different variants of hGCNF have been reported to date, which differ only in the structure of the amino-terminal domain of the protein. The hGCNF variant reported here (AF 004291) has two deletions (dots and squiggles) compared to the hGCNF variants hsu64876 and hsu80802, but has a common deletion (dots) with the hGCNF variant s83309.

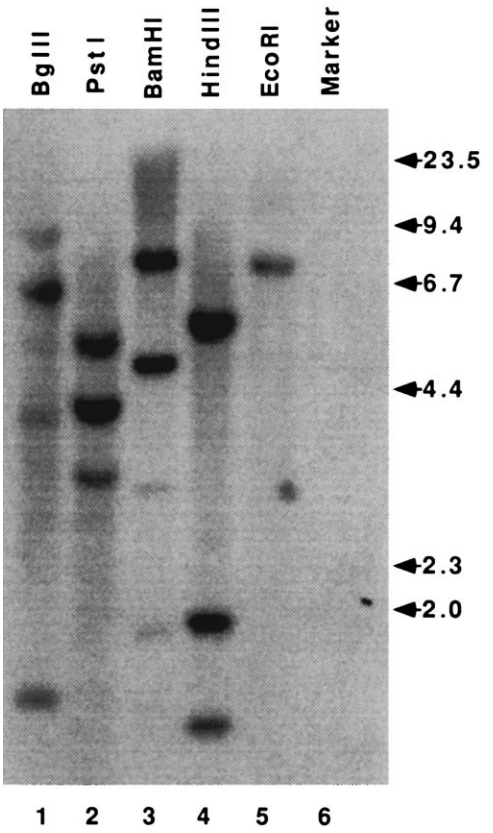

Fig. 2. Southern blot analysis of the hGCNF gene. A human GENO-BLOT (Clontech) was probed with the hGCNF cDNA at high stringency. The size of the molecular weight marker bands is indicated on the right.
←

*Eco*RI digested DNA and two bands with *Bam*HI digested DNA, which is consistent with the restriction map of the hGCNF cDNA (Fig. 2). These results indicate that GCNF is most likely a unique gene. No substantial homology to any other sequences was apparent from the Southern blot analysis. It exhibits several additional weaker bands in the lanes of the DNA digested with *Bam*HI and *Bgl*II. The mGCNF gene has several very small exons (Cooney and Katz, unpublished results), therefore the weak bands on the Southern blot might be a result of the hybridization to the very small exons of the hGCNF gene.

In order to determine on which human chromosome the hGCNF gene is localized and the specific locus within that chromosome, chromosomal localization was employed. Using FISH analysis the GCNF gene was localized to chromosome 9 at the q33–34.1 locus (Fig. 3). At present, of the inherited diseases that have been located to that region of chromosome 9 none of the disease etiologies fits with what is known of the biology of GCNF.

3.3. Analysis of the expression pattern of hGCNF

By Northern blot analysis several groups have demon-

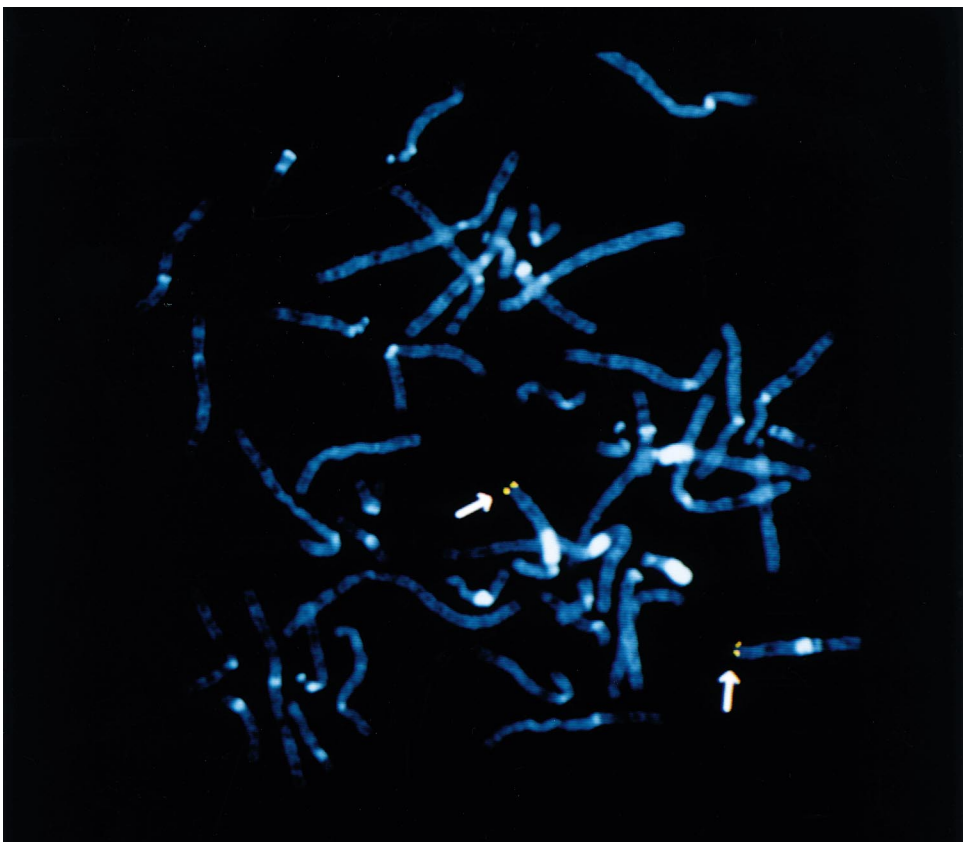

Fig. 3. Chromosomal localization of the hGCNF gene. A BAC clone, obtained after PCR screening of a human genomic DNA library, was non-isotopically labeled and used as a probe in the fluorescent in situ hybridization. The chromosomes were counterstained with DAPI. Images were obtained using cool-CCD camera. The yellow signal (arrows) represents a position of the gene on the chromosome. It is apparent on a single chromosomal pair (chromosome 9) between the bands q33 and q34.1.

strated high level of hGCNF expression in testis and some expression in the ovary, colon and prostate [31,32]. We have observed similar results. To determine the specific cells within the testis in which the GCNF message is localized we employed in situ hybridization. Both sense and antisense riboprobes were generated from the hNF5 clone. No significant

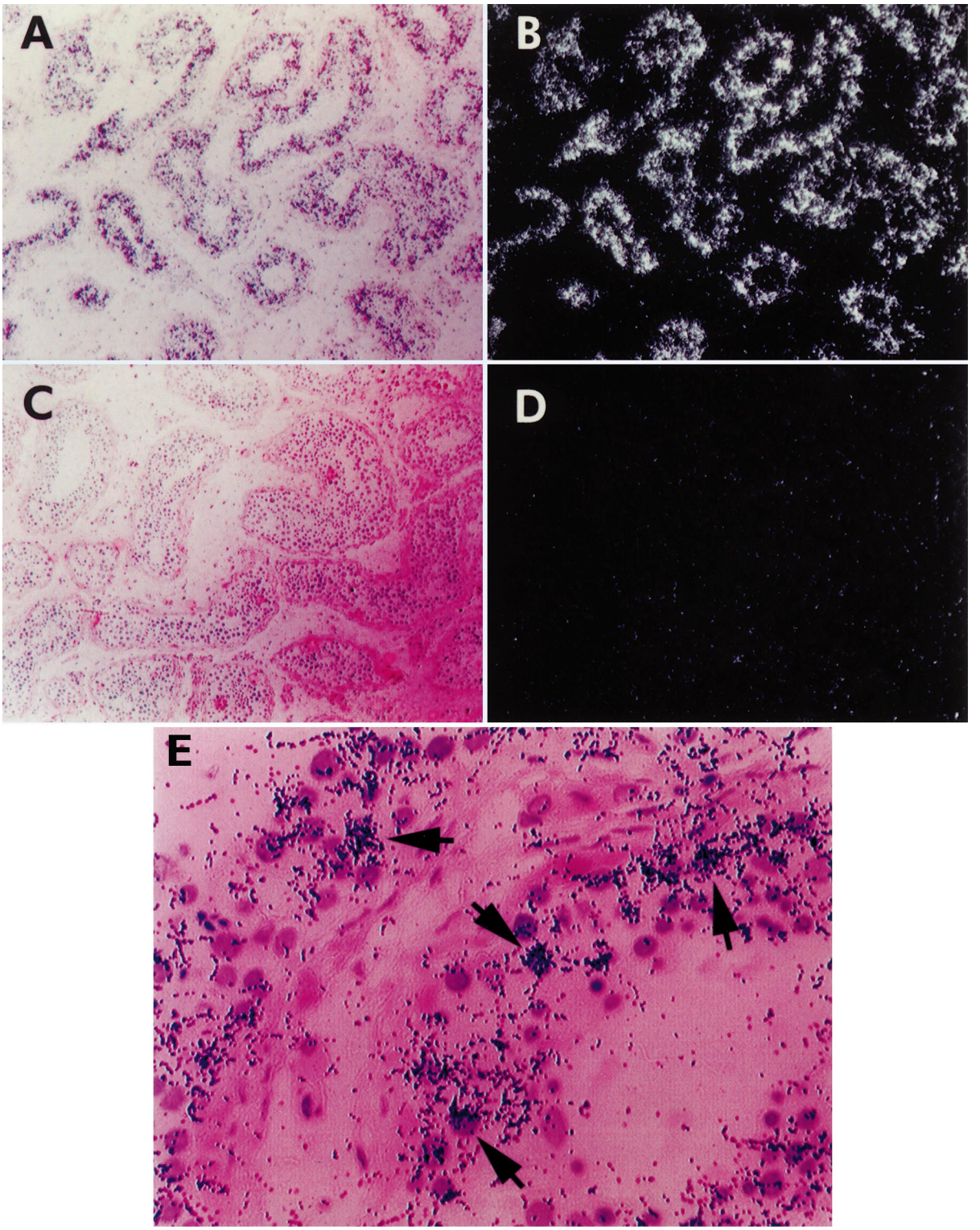

Fig. 4. In situ hybridization analysis of the expression of hGCNF in human testis. A: Light field of the human testis histology stained with hematoxylin and eosin. B: Dark field of the section shown in A probed with a P-33 antisense riboprobe to hGCNF. A strong signal (white silver grains) is clearly detected within the seminiferous tubules. C: Light field of the human testis histology stained with hematoxylin and eosin. D: Dark field of the section shown in C probed with a P-33 sense riboprobe to hGCNF. Only backgound signal is detected. E: High magnification of the light field of a testis section probed with P-33 antisense riboprobe to GCNF (A). A strong signal (silver grains, black dots) was detected in the spermatogenic cells (arrows).

signal above background was observed in the in situ hybridization of the human testis with the hGCNF sense probe (Fig. 4D). In contrast a strong and specific signal, with the antisense probe to hGCNF, was observed in the seminiferous tubules of the human testis (Fig. 4B). The expression of the hGCNF gene was restricted to spermatogenic cells (Fig. 4E). In spermatogenic cells the expression of GCNF appears to be strongest in late spermatocytes (Fig. 4E, arrows) with less signal over the spermatids. Thus, in the testis the hGCNF gene is expressed in a germ cell specific of manner in round spermatids and late stage spermatocytes.

4. Discussion

As previously described mGCNF is a germ cell specific orphan receptor in the adult mouse. In male germ cells it is expressed postmeiotically in round spermatids. In oocytes its expression is detected mid meiosis in growing follicles [17,26]. Here we report the cloning of a novel human GCNF homolog, analysis of its expression by the in situ hybridization, chromosomal localization.

By the fluorescent in situ hybridization we have mapped hGCNF to chromosome 9 locus q33–34.1 and mouse GCNF to chromosome 2 (Cooney, unpublished data). Comprehensive linkage analysis revealed no human or mouse abnormalities of the tissues in which GCNF was detected. Southern blot analysis indicates that there are no closely related members in that subgroup of the orphan receptors. A combination of the results of the linkage analysis, the extremely high homology among evolutionarily distant species, and the fact that GCNF null mutant mice die prenatally (Cooney and Katz, personal communication) suggest an important non-redundant function of GCNF in development.

Chromosomal FISH experiments resulted in a single localization site. In addition, the fact that the DBDs and LBDs are identical in all hGCNF variants indicates that they do not represent orphan receptor isoforms encoded by separate genes, but rather are probably a result of a differential intron/exon splicing of pre-mRNA derived from a single hGCNF gene.

Four reported isoforms provide evidence for alternate exon splicing in the N-terminal part of the human homolog (Fig. 1). At least three different variants have been cloned in other laboratories. The hGCNF cDNA we have cloned lacks 22 amino acids in the N-terminus region relative to the other three variants. It also has a four amino acid deletion found in one of the other variant (S83309). The functional significance of the N-terminal variations are at present is unknown. There is evidence that the N-terminus of GCNF modulates DNA binding: in the absence of the N-terminus GCNF can bind DNA as a monomer whereas a intact receptor binds DNA only as a dimer (Cooney, personal communication). Multiple N-terminal variants of hGCNF may therefore have a different DNA binding characteristics.

The in situ hybridization studies showed that like in mouse the expression of hGCNF in testis is restricted to the germ cells. In contrast to rodents, the expression of hGCNF is not restricted to post-meiotic germ cells [23,26]. In human germ cells GCNF is expressed pre-meiotically in spermatocytes. In fact the in situ analysis seems to indicate that the highest levels of GCNF expression is in the late stage spermatocytes. In the mouse and rat there is a peak of expression of GCNF in round spermatids at stages VII and VII, respectively. These stages are some of the last round spermatid stages prior to elongation. The difference in the temporal expression in spermatogenesis of the human homolog versus the rodents is probably due to underlying differences in spermatogenesis in these species.

In the gonads, exclusive germ cell expression of the GCNF in the developing gametes allows us to postulate that GCNF is an important factor in spermatogenic maturation. Due to the similarity of the GCNF pattern of expression in men and mice, the mouse is a good model to study its role in the spermatogenesis and fertility. To determine this cell specific knock outs of GCNF in oocytes and spermatogenic cells are being generated. These studies will show whether or not fertility could be modified by targeting the GCNF signaling pathway. To address the possible role of GCNF in the abnormal germ cell development the study of GCNF expression in various germ cell tumors is now under way.

References

[1] Evans, R.M. (1988) Science 240, 889–895.
[2] Kastner, P., Mark, M. and Chambon, P. (1995) Cell 83, 859–869.
[3] Mangelsdorf, D.J. and Evans, R.M. (1995) Cell 83, 841–850.
[4] Conneely, O.M. and O'Malley, B.W. (1994) in: Molecular Biology Intelligence Unit: Mechanism of Steroid Hormone Regulation of Gene Transcription (Tsai, M.-J. and O'Malley, B.W., Eds.), pp. 111-113. R.G. Landes, Austin, TX.
[5] Laudet, V., Hanni, C., Coll, J., Catzeflis, F. and Stehelin, D. (1992) EMBO J. 11, 1003–1013.
[6] Laudet, V. and Adelmant, G. (1995) Curr. Biol. 5, 124–127.
[7] O'Malley, B.W. and Conneely, O.M. (1992) Mol. Endocrinol. 6, 1359–1361.
[8] Zanaria, E., Muscatelli, F., Bardoni, B., Strom, T.M., Guioli, S., Guo, W., Lalli, E., Moser, C., Walker, A.P., McCabe, E.R., Meltinger, T., Monaco, A.P., Sassone-Corsi, P. and Camerino, G. (1994) Nature 372, 635–641.
[9] Thummel, C.S. (1995) Cell 83, 835–839.
[10] Willy, P.J. and Mangelsdorf, D.J. (1997) Genes Dev. 11, 289–298.
[11] Luo, X., Ikeda, Y. and Parker, K.L. (1994) Cell 77, 481–490.
[12] Luo, X., Ikeda, Y., Schlosser, D.A. and Parker, K.L. (1995) Mol. Endocrinol. 9, 1233–1239.
[13] Sadovsky, Y., Crawford, P.A., Woodson, K.G., Polish, J.A., Clements, M.A., Tourtellotte, L.M., Simburger, K. and Milbrandt, J. (1995) Proc. Natl. Acad. Sci. USA 92, 10939–10943.
[14] Shinoda, K., Lei, H., Yoshii, H., Nomura, M., Nagano, M., Shiba, H., Sasaki, H., Osawa, Y., Ninomiya, Y., Niwa, O., Morohashi, K.I. and Li, E. (1995) Dev. Dynamics 204, 22–29.
[15] Muskatelli, F., Strom, T.M., Walker, A.P., Zanaria, E., Recan, D., Meindl, A., Bardoni, B., Guioli, S., Zehetner, G., Rabl, W., Schwarc, H.P., Kaplan, J.-C., Camerino, G., Meltinger, T. and Monaco, A.P. (1994) Nature 372, 672–676.
[16] Duncan, S.A., Manova, K., Chen, W.S., Hoodless, P., Weinstein, D.C., Bachvarova, R.F. and Darnell Jr., J.E. (1994) Proc. Natl. Acad. Sci. USA 91, 7598–7602.
[17] Chen, W.S., Manova, K., Weinstein, D.C., Duncan, S.A., Plump, A.S., Prezioso, V.R., Bachvarova, R.F. and Darnell, J.E. (1994) Genes Dev. 8, 2466–2477.
[18] Cooney, A.J., Leng, X., Tsai, S.Y., O'Malley, B.W. and Tsai, M.-J. (1993) J. Biol. Chem. 268, 4152–4160.
[19] Cooney, A.J., Tsai, S.Y., O'Malley, B.W. and Tsai, M.-J. (1992) Mol. Cell. Biol. 12, 4153–4163.
[20] Burbach, J.P.H., da Silva, S.L., Adan, R.A.H., Cooney, A.J., Tsai, M.-J. and Tsai, S.Y. (1994) J. Biol. Chem. 269, 1–8.
[21] Leng, X., Cooney, A.J., Tsai, S.Y. and Tsai, M.-J. (1996) Mol. Cell. Biol. 16, 2332–2340.
[22] Tsai, S.Y. and Tsai, M.-J. (1997) Endocr. Rev. 18, (2) 229–240.
[23] Chen, F., Cooney, A.J., Wang, Y., Law, S.W. and O'Malley, B.W. (1994) Mol. Endocrinol. 8, 1434–1444.

[24] Hirose, T., O'Brien, D.A. and Jetten, A.M. (1995) Gene 152, 247–251.
[25] Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P. and Evans, R.M. (1995) Cell 83, 835–839.
[26] Katz, D., Neiderberger, C., Slaughter, G.R. and Cooney, A.J. (1997) Endocrinology 138, (10) 4364–4372.
[27] Süsens, U. and Borgmeyer, U. (1996) Biochim. Biophys. Acta 1309, 179–182.
[28] Lei, W., Hirose, T., Zang, L.-X., Adachi, H., Spinella, M.J., Dmitrovsky, E. and Jetten, A.M. (1997) J. Mol. Endocrinol. 18, 167–176.
[29] Kleene, K.C. (1996) Mol. Reprod. Dev. 13, 268–281.
[30] Baldini, A. and Lindsau, E.A. (1994) in: In Situ Hybridization Protocols (Choo, A.K.H., Ed.), Vol. 33, pp. 75–83, Humana Press, Totowa, NJ.
[31] Kapelle, M., Kratzschmar, J., Huseman, M. and Schleuning, W.-D. (1997) Biochim. Biophys. Acta 1352, 13–17.
[32] Joos, T.O., David, R. and Dreyer, C. (1996) Mech. Dev. 60, 45–57.