

Full Length Research Paper

Structural and functional analysis of the promoters of the human glycerol kinase gene family

Banu Az-Zubair, M. K.

Department of Molecular Biology and Bioinformatics, National Biotechnology Development Agency, Area 11, Garki-Abuja, Nigeria. E-mail: azzubair@nabda.gov.ng. Tel: +234 9 671 5691 or +234 9 314 5473.

Accepted 9 November, 2007

During spermatogenesis the X-chromosome is inactivated, but several glycolytic cycle enzymes have been found to have spermatogenic cell specific isoforms encoded by the X-chromosome counterpart genes located on the autosomes. Here, the promoters of the human glycerol kinase gene family were cloned: the somatic gene called *GK-Xp*, and the two testis-specific isoforms called *GK-1* and *GK-2*, using *Alu*-GSP primers. The structure and function of the promoters were determined by sequencing and in situ expression studies using GFP as reporter. The transcription initiation sites were mapped by RACE and, these were found at -112 for *GK-Xp*, -46 for *GK-1*, and for *GK-2* at -57 nucleotide positions, upstream of the translation start codon. The *GK-Xp* promoter specifically contains a TATA-like motif at -26, and a CCAAT motif at -149, which are not found in either the *GK-1* or *GK-2* promoters. Interestingly, *GK-1* contains an Inr motif, a Sp1 binding site at -11, and a DPE motif at +27 positions, whereas *GK-2* contains an Inr motif, a CRE at -11, a DPE at +29 and a RARE motif at +43 positions. In addition, they both contain MEP-2 and Ets motifs, in common with other testis-specific metabolic enzymes genes, such as *PGK-2* and *PDHA-2*.

Key words: Spermatogenesis, transcription, gene expression.

INTRODUCTION

Glycerol kinase (GK) is a member of a small group of kinases termed *ambiquitous* enzymes, meaning "both places", that are found in the cytosol, or membrane-bound, associated with the voltage-dependent anion channel of the mitochondrial outer membrane (Wilson, 1978). Glycerol kinase (ATP: glycerol 3-phosphotransferase, EC 2.7.1.30) is involved in the metabolism of endogenously derived and dietary glycerol by the phosphorylation of glycerol to glycerol 3-phosphate (McCabe, 1983). This reaction has been studied in various organisms, including bacteria and protozoa (Newsholme, 1969; Pettigrew, 1988). In mammals, GK activity has mainly been identified in liver and kidney and, at low levels, in intestinal mucosa, adipose tissue, cardiac and skeletal tissues (McCabe, 1989). The gene encoding human glycerol kinase expressed in somatic tissues was isolated by several groups, and has been mapped to the short arm of the X-chromosome (Guo, 1993; Walker et al., 1993; Sargent et al., 1994). However, the X-chromosome is inactivated at specific stages during spermatogenesis. Thus, Northern hybridisation analyses of human male

adult tissues and clones isolated from a human testis cDNA library have revealed at least two testis-specific glycerol kinase genes. Both of these testis-specific transcripts mapped to chromosome 4 (Sargent et al., 1994). The mouse homologue of these genes has recently, also, been identified (Pan et al., 1999). However, little is known about the GK regulatory elements in any species.

Spermatogenesis is the series of processes that produces the male gamete or spermatozoa. It is an exceedingly complex and highly regulated developmental process that begins with the division and differentiation of the spermatogonial stem cell within the seminiferous tubules of the testis, and ends with spermiogenesis, the process that transforms round spermatids into mature sperm. During this process a large variety of genes normally expressed in somatic tissue are repressed following germ cells X inactivation, but their isologues on the autosomal chromosomes are specifically transcribed in the testis, while others are expressed in both spermatogenic and somatic tissues. Repeated cDNA analyses of

both somatic and testis tissues have shown that the *GK* genes are differentially expressed. *GK-Xp* is predominantly expressed in somatic tissues, such as kidney and liver, whereas *GK-1* and *GK-2* on chromosome 4, are specifically expressed in the testis (Sargent et al., 1994), a result expected because of X-inactivation (McCarrey et al., 1996; Erickson, 1993).

Gene expression is primarily regulated at the transcriptional level and the core promoter architecture plays a most crucial role. The transcription start site (TSS) and transcription factor binding motifs identify the 5' flanking region of human genes. Transcription start sites may be determined by rapid amplification of cDNA ends, RACE, while general transcriptional factor binding motifs may be determined on the basis of recently published sequences and the TRANSFAC database. Although the transcription start site generally marks the end of a promoter (exception include PolIII promoters), critically important motifs are sometimes located further downstream in the 5' UTR. Nevertheless, generally, two different segments of promoter structure and function have been described for RNA polymerase II transcription in eukaryotes: a core promoter region of around 50 – 40 (Roeder, 1996) nucleotides adjacent to the transcriptional start site and more distant promoter elements (Roeder 1996). In RNA polymerase II transcription core promoter elements are defined as 'minimal DNA elements that are necessary and sufficient for accurate transcription in a reconstituted cell-free system'. A core promoter contains two key genetic elements, which can function independently or synergistically: the TATA element (consensus TATA^A/_TA^A/_T) located upstream of the transcription start site near -30 to -25 and/or an initiator (Inr) element, a pyrimidine-rich sequence (consensus PyPyA₁N^T/_APyPy), located around the transcription start site (Smale and Baltimore, 1989). To investigate whether the differential transcription of the glycerol kinase genes, between somatic and testicular tissues, is based on core promoter structure difference, their promoters were cloned, their structures established by sequencing and functionally tested for differential activity in testis germ cells and non-germ cells, using GFP as reporter. The structures of the core promoter regions were found to be different between the somatic *GK-Xp* and testis isoforms (*GK-1* and *GK-2*). Although the *GK-1* and *GK-2* core promoters share similar transcription motifs, they were also found to be fundamentally different.

MATERIALS AND METHOD

PCR

PCR was done using 20 pmol of primer pair, 1.5 mM Mg⁺⁺, 2.5 mM dNTPs, 1X Taq polymerase buffer, 1 - 2 units of Taq polymerase enzyme, 10 – 50 ng template DNA in 50 µl reaction volume made up with water. ExpandTM Long Template PCR System was used according to the manufacturer's protocols. Amplification was run for 40 cycles (94°C, 1min; 55 - 72°C, 10 s to 1 min [depending on the primers]; 72°C, 2 min).

PAC library screening and isolation of genomic clones

A human PAC library (Human Genome Mapping Project (HGMP) Resource Centre, Hinxton, Cambridge) constructed using a normal male blood donor was screened by PCR, using gene specific primers, to isolate genomic clones containing the testis-specific genes *GK-1* and *GK-2* (Table 1). To isolate the promoter regions of the human *GK* genes a human PAC library was screened by PCR using gene-specific primers encompassing about 300 bp of the coding region (data not shown). The primers *GK-p2* and *GK-p4* were used to screen the library as these were expected to identify the *GK* gene locus on Xq, Chromosome 1, and at least one of the two loci on Chromosome 4, but not the *Xp* locus. Ten positive clones were isolated and analysed further by amplification with primers *GK-p15* and *GK-p31*, which generated a product covering the entire open reading frame (ORF) followed by sequencing (data not shown) and digestion with *Taq* I and *EcoR* I restriction enzymes (data not shown). The *EcoR* I and *Taq* I digestion was to confirm the identity of the clones, because of differences between loci. The following four clones were obtained using *GK-2* gene specific primers: 63N19; 121O20; 186G15 and 220I13. Seven clones were isolated with the *GK-1* specific primers: 54H13; 54H14; 56F13; 72F20; 77J23; 155M12 and 155M14. The *GK-Xp* genomic clone, 7E22 was donated by Dr. C. A. Sargent.

Primers *GK-p163* and *GK-p1090*, specific for the *GK-2* gene also on chromosome 4, were used to generate a product of about 1kb (data not shown). Four clones were isolated, three of which were further analysed by amplification of the ORF using primers *GK-pA13* 3' and *GK-p31*, followed by sequencing (data not shown) and digestion with *Taq* I and *EcoR* I restriction enzymes (data not shown).

Cloning the 5' flanking regions of the *GK* gene family

Gene specific primers, positioned within the first 200 bp of the available coding regions was paired with *Alu* primers to amplify the 5' flanking regions, using ExpandTM Long Template PCR System according to the manufacturer's protocols (Boehringer Mannheim). The first PCR products were diluted 1/100 in distilled water and 5 µl used for nested PCR in the same conditions. Both the first and nested PCR products were examined on 1% agarose gel stained with ethidium bromide and visualised by UV in the MultimageTM Light Cabinet (Alpha Innotech Corporation) and analysed with Chemi-Imager software and camera (Alpha Innotech Corporation). The nested PCR products were directly sub-cloned into pGEM-T easy vectors (Promega).

DNA sequencing

Sequencing was performed using the dyedeoxy termination method using Prism ready reaction terminator cycle sequencing kit (Applied Biosystems) on a Perkin Elmer Applied Biosystems Automated Sequencer, ABI 373A, and analyses were done with the manufacturer's installed software. Briefly, 500 ng of DNA template, 2.0 pmol of primer, and 9.5 µl of premix in 20 µl total volume were incubate for 30 cycles at 96°C for 30s, 50°C 15s and 60°C for 4 min. The product was purified and precipitated in ethanol and run on 6% denaturing polyacrylamide gel for at least 16 h.

Mapping of transcription start site

The transcription start site was determined with Rapid amplification of cDNA ends (RACE), using Clontech's human testis cDNA kit (Marathon-ReadyTM cDNA, cat. No. 7414-1) according to the manufacturer's instructions. Briefly, internal gene-specific primers

Table 1. The oligonucleotide primers sequence.

Primer	5'—Sequence→3'	Tm°C	Use
GK-p2 (F)	cgc tgg cta aga gac aat ctt gg	55	PCR + Seq.
GK-p4 (R)	cta atg cag caa aag caa tat ggc	55	PCR + Seq.
GK-p19 (F)	ccc cct cca ttg cca taa cat c	55	PCR
GK-p19 (R)	gat gtt atg gca atg gag ggg	55	PCR(<i>Alu</i>)
GK-p31 (F)	aag ctg gtt tca tgg cag cc	58	PCR
GK-p32 (R)	gtc cac cgc ccc cac caa t	61	PCR + Seq.
GK-p33 (R)	ctc gtc acc tgc ccc tcc c	62	PCR
A13-3' (R)	aca agt cta ggg ttt tca tgg g	55	PCR
A13-p163 (F)	aaa gaa gga tgg gtg gaa caa	55	PCR (PAC)
A13-p1090 (R)	aaa ggc tgg gac aaa gta aca	55	PCR (PAC)
142-p170 (R)	ttc cac cca tcc ttc tct tg	58	PCR (<i>Alu</i>)
142-p39 (R)	ccc caa aac tgc ctt ctt	54	N-PCR + Seq
A13-p162 (R)	ttc cac cca tcc ttc ttt tg	58	PCR(<i>Alu</i>)
A13-p78 (R)	aga aag cga gtg gag ttg g	56	N-PCR + Seq.
<i>Alu</i> -1	tcc caa agt gct ggg att aca g	55	PCR
<i>Alu</i> -2	ctg cac tcc agc ctg gg	55	PCR
<i>Alu</i> -3	gat cgc gcc act gca ctc c	55	PCR
<i>Alu</i> -4	gga tta cag gcg tga gcc ac	55	PCR
Xp2-5'	ccc caa aac tgc ctt ctt tga ggc t	72	5'-RACE
Xp1-5'-N	ctt tga ggc tgc cat gaa acc agc tt	72	5'-RACE +Seq.
142b-5'	ggc tct cca gtt atc ttg tcc cag gct a	72	5'-RACE
142a-5'-N	cac cca tcc ttc tct tgg gaa ctc ttg t	72	5'-RACE +Seq.
A13c-5'	tca agc cac acc aca gca ttg tag aga g	72	5'-RACE
A13b-5'-N1	ggt cgt caa gtt tct cac acg ttc tcg c	72	5'-RACE+Seq
A13a-5'-N2	ggt ctt gtt cca ccc atc ctt ctt ttg g	72	5'-RACE+Seq

Nucleotide sequences of primers. The names, Tm temperature and the purpose for which the primers were used are indicated.

GSP complementary were paired with an adaptor primer (AP1) to amplify the 5' flanking region of the cDNA. The first PCR products were diluted 1/100 in distilled water and 5 µl used for nested PCR in the same conditions, using nested GSP and N-AP2 primers. The products were sub-cloned into pGEM-T easy vectors.

Tissue culture

A human male liver adenocarcinoma (Hep-G2) was obtained from American Type Culture Collection (ATCC). Joanne Davis and Dr. Paul Edwards (Department of Pathology, University of Cambridge) donated a human mammary cell adenocarcinoma (MCF-7) cell line. Hep G2 cells were maintained in Minimum Essential Medium (MEM) supplemented with 1% Penn-Strep, 1% glutamine, 1% Non-essential amino acids (NEAA), and 10% Foetal Calf Serum (FCS), at 37°C, 5% CO₂. MCF-7 cells were maintained in DMEM, supplemented with 1% Penn-Strep, 1% glutamine, 1% insulin and 10% FCS, at 37°C, 5% CO₂.

Mouse spermatogonial/spermatocytes (GC1) (CRL-2053) and spermatocytes/round spermatids (GC2) (CRL-2196) cells were also obtained from ATCC. The cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 1% Penn-Strep, 1% sodium pyruvate, and 10% Foetal Calf Serum (FCS), at 37°C, 5% CO₂.

Deletion constructs

Three different fragments of the *GK-Xp*, *GK-1* and *GK-2* 5' flanking regions were functionally tested for promoter activity. The *GK-Xp* 5 kb and 2 kb fragments were individually generated from *Alu*-GSP PCR, whereas the 300 bp fragment was generated by deletion of the cloned 2 kb fragment, with *Apa* I restriction enzyme, and re-ligating the remaining fragment in the expression cassette. Similarly, the *GK-1* 5 kb and 1.6 kb fragments were separately generated by *Alu*-GSP PCR, whereas the 1 kb fragment was generated with a double digestion of the 1.6 kb fragment with *EcoR* I and *Bgl* II enzymes, and re-ligating into the expression cassette. Finally, the *GK-2* 5 kb and 1.7 kb fragments were also separately generated with *Alu*-GSP PCR. However, the 1 kb fragment was generated from the 1.7 kb fragments with *Hind* III restriction enzymes, the 1 kb fragment isolated and sub-cloned into the GFP expression vector. All the constructs include the 5' UTR of the gene, ending just upstream of the translation start site.

Transfection and GFP assay

DNA was prepared for transfection by purification through Qiagen column, according to the manufacturer's protocols. A day before transfection 1 x 10⁵ cells was seeded onto a sterile glass cover slip

in each well of a 6-well plate. 1 µg of each construct was co-transfected with equal amount of pGL-3 luciferase vector as internal control in triplicate, with Effectene™ as transfecting agent, according to the manufacturer's protocols. Cells were generally prepared for reporter assays after 48 hours post-transfection. Culture media was removed from each well and the cover-slips transferred into fresh labelled 6-well plates and washed three times with 1X PBS, after which they were layered with 2 ml PBS/4% paraformaldehyde for 30 min at room temperature. The PBS/4% paraformaldehyde was then removed and the cells washed twice in 1X PBS. The glass cover slips were removed and mounted on a glass slide for the analysis of the cells on a fluorescence microscope. The remaining cells in the culturing plates were used for the luciferase assays. Briefly, after the removal of culture media, the cells are washed once in 1x PBS. The PBS was aspirated and previously prepared 250 µl 1x lysis buffer was added per well, and the plates left at room temperature for 10 mins. Cells were dislodge by scraping after the 10 min incubation and transferred into 1.5 ml eppendorf tubes, and centrifuged for 10 sec, at 12 000 x g. The supernatant was transferred into fresh eppendorf tubes for analysis.

To normalise the reporter gene activity, equal amounts of each GFP-construct cassette and luciferase pGL3 internal control were transfected into each cell type in triplicates as stated above. 48 h after transfections the cells were lysed and the supernatant assayed in triplicate using a luminometer. Means of the triplicate luciferase activity for each construct were calculated and the background luciferase activity subtracted from each experiment. The mean net luciferase activity from each transfection was then normalised to the net mean value of the GFP fluorescence activity.

RESULTS

Determination of transcription start site

To characterise the 5' flanking region, the transcription start site was determined by RACE. The 5' RACE assays provided extended single band products that indicated that there was only one major transcription initiation site for each of the GK genes (Figure 4). Following determination of the transcription start sites, fragments immediately upstream of the translation start site were cloned by nested PCR from the genomic clones. Analysis of the sequence data show that the 5' flanking regions of *GK-1* and *GK-2* have no TATA box, while *GK-Xp* has a TATA containing promoter; thus the *GK-1* and *GK-2* have TATA-less promoters. Furthermore, the promoter regions show several predicted regulatory sites and, within 500bp of the core promoter, the *GK-Xp* contains a TATA box at -27, a CCAAT box at -149 positions but lacks a CRE motif, whereas *GK-1* and *GK-2* contain Inr, DPE and CRE motifs and, significantly lack a CCAAT box (Figures 1A-C).

Comparison of the GK gene family core promoters motifs

The key motifs in the GK gene family promoters were analysed, indicating a strong similarity between the *GK-1* and *GK-2* promoters (Figure 2). The difference between *GK-Xp* and the testis-specific *GK-1* and *GK-2* is charac-

teristic of the difference between most testis-specific genes and general promoters active in somatic tissues.

GK-Xp Promoter

Sequence analysis of the *GK* gene family core promoters revealed that the *GK-Xp* promoter contains a TATA-like motif (sequence GGTAAG) at -27, and CCAAT motif at -150 positions. The putative TATA-like motif of the *GK-Xp* has only slight similarity with the consensus TATA box in higher eukaryotes, however, it is similar to the TATA box found in lower eukaryotes, such as *Entamoeba histolytica* (sequence GTATTTAAA^{G/C}) (Singh et al., 1997). Furthermore, the CCAAT motif sequence ATTGG, (consensus CCAAT) is reversed, in 3' to 5' orientation. Interestingly, the *GK-Xp* promoter contains two glucocorticoid response elements, a positive regulatory motif, at -415, and a negative regulatory motif, at -275. However, the sequence motif to which the glucocorticoid regulatory factor binds to induce positive regulation (GRE) is different from the motif it binds to effect negative regulation (nGRE) (Sakai, 1988) (Churikov et al., 2004). Nonetheless, there is some similarity between the two sequences. The positive regulating glucocorticoid response element, GRE, (consensus: GGT ACANNNTGTTCT) in *GK-Xp* (sequence: GGAACAAGCAGTTCT) is remarkably similar to the consensus motif. The nucleotides in bold represent the consensus 'core' GRE binding site, while the 'core' and the 5' sequence flanking the motif is critical for regulation by the transcriptional factor. The nucleotide positions in *Italics* within the *GK-Xp* sequence represent possible polymorphisms. The negative regulating glucocorticoid response element, nGRE, is at -275, further downstream from the positive regulating motif, at -415 (consensus: ATYACNNNNNTGATCW) and the nGRE in *GK-Xp* (sequence: ATGGCTCTGCTGTCCG).

GK-1/GK-2

In contrast to *GK-Xp*, both *GK-1* and *GK-2* promoters contain the initiator motif at the transcription start site (Figures 1B and C) (*GK-1*, sequence TCACCCA; *GK-2*, sequence CCAACAC). The most highly conserved nucleotides in the initiator element of greater than 500 eukaryotic genes appear to be -1(C), +1(A), and +3(T) (Bucher, 1990). The initiator motif, also referred to as a YY1 motif, has been shown through mutational analysis to consist of a core sequence of 6 base pairs CCATN (T/A) (Ye et al., 1999). The *GK-1* 'core' promoter region lacks both a CCAAT and TATA boxes. In addition, it has a cyclic-AMP response element (CRE) motif at position -28. It has a Sp1 site directly adjacent to the transcription start site, at -6 position, and HRE direct repeats, like the *GK-Xp* promoter, at -10 to -27 positions. Crucially, the positioning of the HRE repeats right next to the transcrip-

GK-Xp

-445 GCACAGCCCA CTGGAGGGAA **CAAGCAG** TC TACACACTTT GAGACCCCTA TGGAGGCGAG
GRE

-385 CCCAGGACA CTAGTGAAG TGCGCTCTTT TCGCCGGCCA GGAGCGCTGC CTGCCAAAGG

-325 **GCAGTACTCG** **CTCCGCC** ATGCCATGCA GCTGGATGGC **TCTGCTGTCC** GCTCGCCGGG
EBP-20 Sp1 nGRE

-265 CCC**CCACGT**G ACCTACTGCG CGGGCCGCCG CCTCCACCGC **GCCCAGGCC**C CGCCCCGTG
USF Sp1

-205 ACGTTCATC GCCCTCTGCC TGTCGAGCG TTCGGCTCTT TTGCGTT**TGCTG**ATTGGCCA
CAAT Box

-145 GCCCAGGTGA CAG**CGCCAC** **CCCCA**CCCCC CCTGACACGC CCCTGGAGCC TTGGGGGTAC
HRE

-85 GCGGTG**GGGG** **GCG**GGGCCTG **GGC**GGAGGG GCGGGGTGAG AAGG**TGCGC** **GCGGGTAAAG**
Sp1 HRE Sp1 TATA Box

-25 **GGGCCG**CCTC GAGCGCGGTC CGAGC**G**TTCA GCGGACGCGC GCGGCCTCGA TCTCTGGACT
Sp1

+36 CGTCACCTGC CCCTCCCCT CCCGCCCG**TCACCCAGGA** AACCGGCCGC AATCGCCGGC
GIM

+95 CGACCTGAAG CTGGTTTCAT G

Figure 1A. Nucleotide sequence of the 5' flanking region of human glycerol kinase, *GK-Xp*, gene. The transcription start site, +1, is shown in bold large font. The TATA box is at -27, and germ cell inhibitory motif (GIM), at +68; both are in boxed bold letters. Hormone response element (HRE), consensus CCC (A/T) CCC, is shown as direct repeats, at -118, and overlapping a Sp1 site at -56 positions, just upstream of the TATA box. A CCAAT box is at -150. There are two types of glucocorticoid response elements (GRE), a negative GRE (nGRE) at -275 and positive GRE at -415 positions.

tion start site, unlike in the *GK-Xp*, suggests that it may be involved in transcriptional activation of *GK-1*. Other important motifs in the core region are the heat shock factor (HSF) at -125 and GATA-1 at -293 positions. The *GK-2*'core' promoter is also characterised with an initiator motif at the transcription start site and, like the *GK-1*, lacks both CCAAT and TATA boxes. Furthermore, like the *GK-1*, the *GK-2* promoter contains a Sp1 site right next to the transcription start site, with an HRE motif adjacently upstream. However, the HRE in *GK-2* is a single motif, not repeats as in *GK-Xp* and *GK-1*. Other motifs common to both *GK-1* and *GK-2* include the GATA and HSF. The *GK-2* also contains a retinoic acid response element (RARE) motif, at +43 position. Interestingly, both *GK-1* and *GK-2* promoters contain a downstream promoter element (DPE) motif, as well as an Ets motif (see Figure 2).

Functional Analysis of the 5' Flanking Regions

To determine the functionality of the promoter sequence,

several fragments of the 5' flanking regions (a 5kb, 2kb and 350bp *GK-Xp*; a 5kb, 1.6kb and 1kb *GK-1*; a 5kb, 1.7kb and 1kb *GK-2* fragments) were cloned into a GFP reporter cassette. Each construct was separately transiently cotransfected with a Luciferase control in triplicates into GC1, GC2, HepG2 and MCF-7 cells. Transcriptional activity driven by each of the *GK* promoters were measured by intensity of the GFP fluorescence, relative to that measured from a promoter-less GFP cassette transfected into the same cells. The histogram of the relative GFP fluorescence is the net activity after subtraction of the promoterless GFP background and normalisation with the luciferase internal control (see M&M above). The cell lines supported transcriptional activities from the individual promoter constructs to varying degrees, while there was no expression of GFP in all the cells from control promoter-less construct. Furthermore, variations in relative transcriptional activities relative from the fragments per cell type will indicate the presence or absence of important transcriptional motifs.

GK-1

Figure 1B. Nucleotide sequence of the 5' flanking region and 5' UTR of testis-specific human glycerol kinase, *GK-1*, gene. The sequence in lower case is part of the coding region. The transcription start site, +1, is in bold large font. There is not TATA box or a CCAAT motif. There is, however, the initiator and several CRE motifs, at -29 and -263 positions, and Sp1 sites, with one next to the TSS. In addition it contains a HRE direct repeats, like the somatic GK, at -11 position. Significantly, it has a downstream promoter element, DPE, at +27 position, MEP-2 at -263 and Ets at -75 positions. The sequence in lower case is part of the coding region.

Promoter Activity in GC1 Cells

The GC-1spg (GC1 cells) represents immortalised pachytene stage primary spermatocytes. The cells were isolated from primary murine pachytene spermatocytes transfected with SV40 Large T antigen ((Hofmann et al., 1992). *GK-Xp* promoter driven transcription in GC1 cells is barely above background level, regardless of the fragment size. This suggests either the presence of inhibitory factors interacting with cognate *cis*-elements in the promoter, or the absence of factors necessary to support transcription from the promoter in GC1 cells. However, both the *GK-1* and *GK-2* promoters were transcriptionally active. The three *GK-2* fragments (5 kb, 1.7 kb and 1 kb) had similar levels of GFP expression, indicating that there

are no significant motifs beyond the 1 kb fragment that will modulate transcription in GC1 cells. The *GK-1* fragments showed greater levels of activity in GC1 cells compared to the *GK-2*; the *GK-1* fragments had about twice as much activity than the *GK-2*. The *GK1* 5 kb fragment had the least promoter activity of the three fragments, which is about the highest activity levels seen in *GK-2* 5' fragments. The *GK-1* 1.6 kb and 1 kb fragments had higher levels of activity, with the 1 kb fragment showing the most activity. The relative similarity in promoter activity between the 1.6 kb and 1 kb fragments suggests that the 1 kb fragment contain the necessary and sufficient elements for efficient transcription in GC1 cells. However, both *GK-1* and *GK-2* 5' flanking regions had greater than x50 relative transcriptional activity in

GK-2

-451 TTTAAAATAA TAGCGCAAAG TTCCACCTGA AACAGCAAGA GAGACGATGG
TAGAGCAACA

Ets

-391 GCTCCAAGGT CCTGAAAAT AAAAAGAAAA AATATGTCAA ATAGAATTTT
ATACCCCGAG

HSF

-331 GAAATACCTT TGAATATTAA TATGAAATAT ATTTTCAGATG TTTAAATTTG
GGAGAATTTA

-271 TTTCCAGCAG ACCTGCATTA TTAAGAAGGG AAATGATAAC GTAGAGATAI
AGATCCACAC

EBP-1**AFP-1****RBP****MEP-2**

-211 AAAAGAGTGA ACGTGCACGC CCGGTAAGT TAATTACATG GATAAATATA
TGATTTAAAA

GATA

-151 AAAAAATTA AATCTTAGAA GACAACTGAG TGGCTGGAGG TGAAGCCTGG
GCTGGAGGGG

GATA**TSF**

-91 TGGGGTGAGA AGGCGCCGCG CCA GTTAAAG GGCA AAGGTT GCCTTGCGCC
TGCGCCTTGT

Ets**CP1****CRE**

-31 GCTCCTGACT GGTTCGCGTT CCCTGCCCC AACACACCGC CGTGGCTTGG
GAAACGGTCC

CRE HRE/Sp1 Inr**RBP**

+29 CAGAGCTGCT GAGCCGACCT ACTGGTGTca tggcagcccc aaagacagca gctgtggccc

DPE**RARE**

Figure 1C. Nucleotide sequence of the 5' flanking region of the testis-specific human glycerol kinase gene, *GK-2*. The transcription start site, +1, is in bold large font. The region contains neither a TATA nor a CCAAT motif. It does contain, however, the initiator element and a couple of CRE and HRE motifs. Interestingly, it contains a DPE, at +29, a RARE motif at +43, and a Sp1 site next to the TSS. It contains an Ets motif at -53, and a MEP-2 site at -212 positions. The sequence in lower case is part of the coding region.

GC1 cells compared to any of the *GK-Xp* 5' flanking sequences.

Promoter activity in GC2 cells

Co-transfection of the SV40 Large T antigen and a temperature-sensitive variant of the p53 gene, under a constitutive promoter control, resulted in the establishment of a second immortalised germ cell line (GC2 cells). Identification of structures resembling proacrosomal granules and observation of haploid peaks, following immuno-histo-chemical and flow cytometric studies

respectively, led the researchers to conclude that the GC2 cells are equivalent to secondary spermatocytes to round spermatid stage cells (Hofmann et al., 1994). Subsequent molecular biological analysis by us, using RT-PCR, also suggests that the GC2 cells contain Pgk-2 and protamine transcripts (data not shown).

There appears to be no difference between GC1 and GC2 cells with respect to *GK-Xp* driven transcriptional activity. Both the *GK-1* and *GK-2* promoter sequences were transcriptionally active in GC2 cells, however, with relatively higher transcriptional activity in GC2 cells than in GC1 cells (Figure 3). Furthermore, the *GK-2* fragments have significantly stronger activity than the *GK-1* in GC2

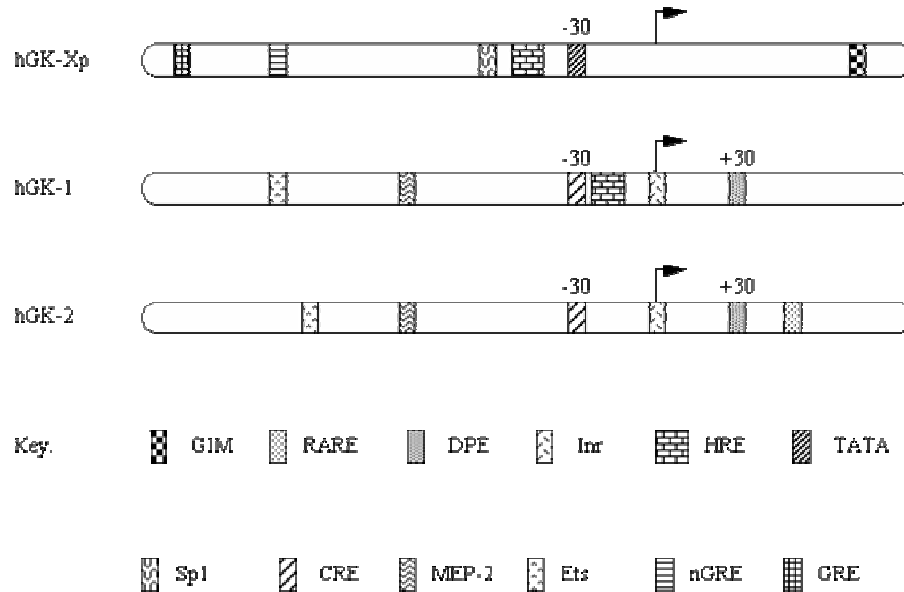


Figure 2. Comparison of the GK gene family core promoter motifs. Note the striking similarity between the GK-1 and GK-2 motifs and their placement within each promoter. The difference between the testis-specific promoters and that of the somatic isologue (GK-Xp) is clearly discernible.

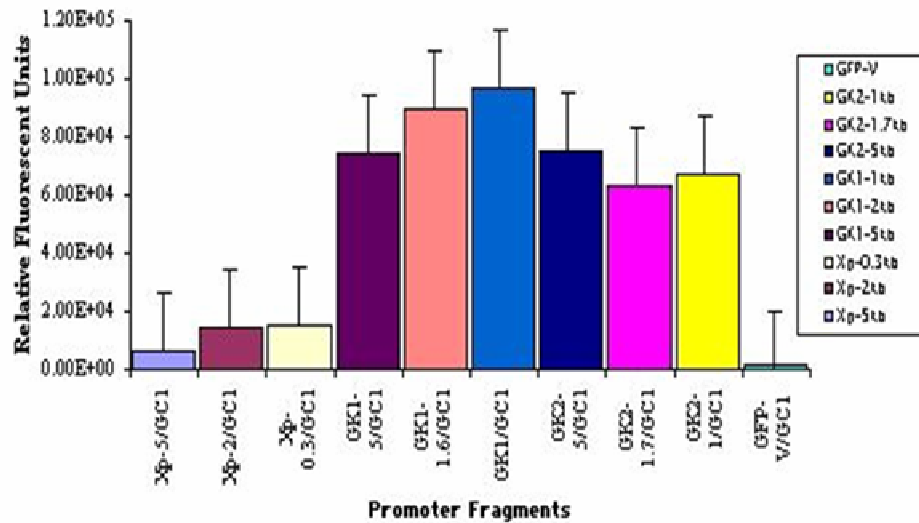


Figure 3A. GK gene family promoters in GCI cells.

cells.

Promoter activity in HepG2 cells

HepG2 is a human male liver adenocarcinoma cell line. The Glycerol kinase gene encoded on the X chromosome is specifically expressed in somatic tissues and, is particularly highly expressed in the liver and kidney. This cell line, thus, served as a positive control for the *GK-Xp*

5' flanking region transcriptional activity. The *GK-Xp* promoter transcriptional activity in HepG2 cells is the reverse of that in the GC1 and GC2 cells, relative to *GK-1* and *GK-2* 5' flanking regions. Although there is a slight transcriptional activity from both *GK-1* and *GK-2* 5' flanking regions in HepG2 cells, they are much lower than that observed from *GK-Xp* approximately an order of a magnitude (see Figure 3). It is interesting to note that *GK-Xp* is most highly expressed in the liver, kidney and muscle, whereas the *GK-1* and *GK-2* are only transcribed

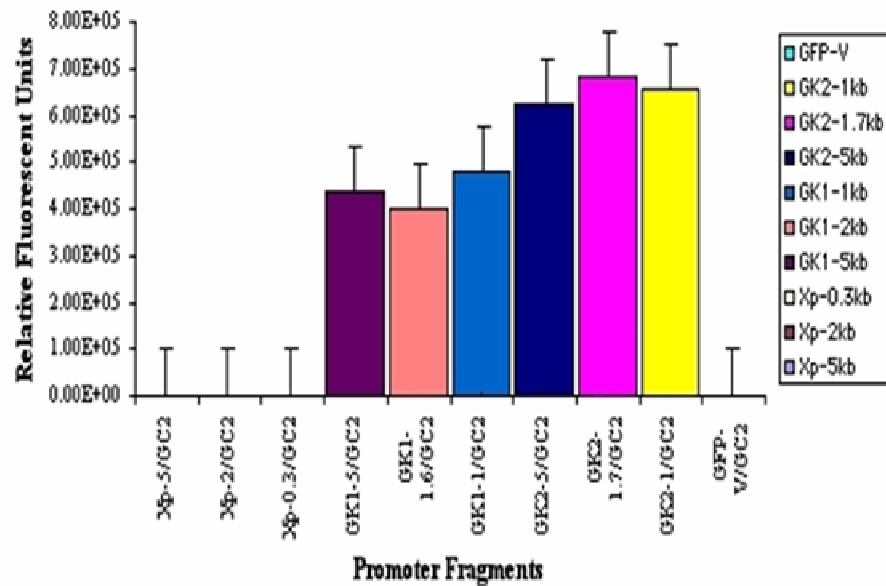


Figure 3B. GK gene family promoters in GC2 cells.

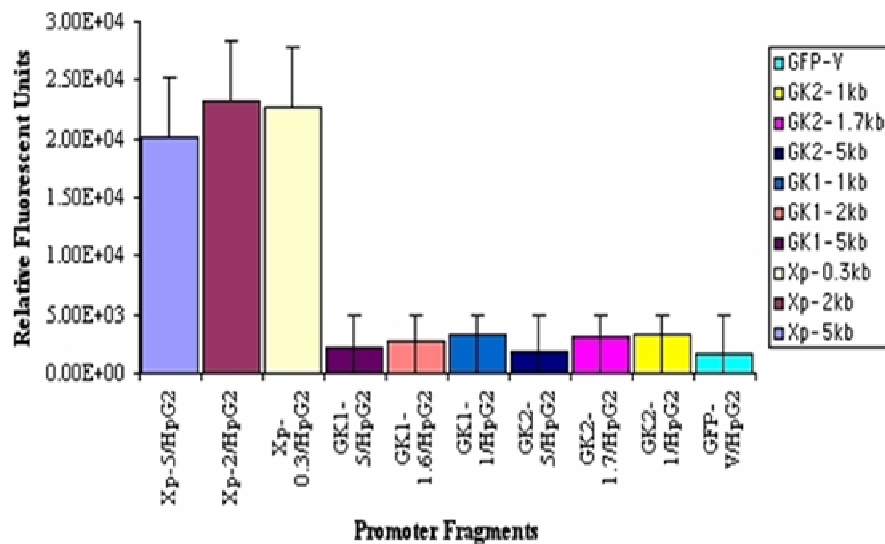


Figure 3C. GK gene family promoters in HepG2 Cells.

in the germ cells. These results, therefore, are entirely consistent with the results observed by Northern blot analysis (Sargent et al., 1994). The result also suggests that the basis of the differential activity probably reside in the 5' flanking sequences and, that the element(s) is contained within the 1 kb fragments of the *GK-1* and *GK-2* genes and the 350 bp of the *GK-Xp* gene.

Promoter activity in MCF-7 cells

MCF-7 cells are human female mammary adenocarcinoma cells. These cells also, like the liver cells, naturally

express the somatic isoform of glycerol kinase, encoded by the X chromosome (McCabe, 1989). The *GK-Xp* fragments showed stronger promoter activity while the *GK-1* and *GK-2* fragments did not show a significant activity (Figure 3). Again, this is consistent with the expression profile of the genes as observed both in vivo and Northern blot analysis.

DISCUSSION

Until fairly recently, there were no immortalised germ cells with which to do *in situ* study of germ cell-specific

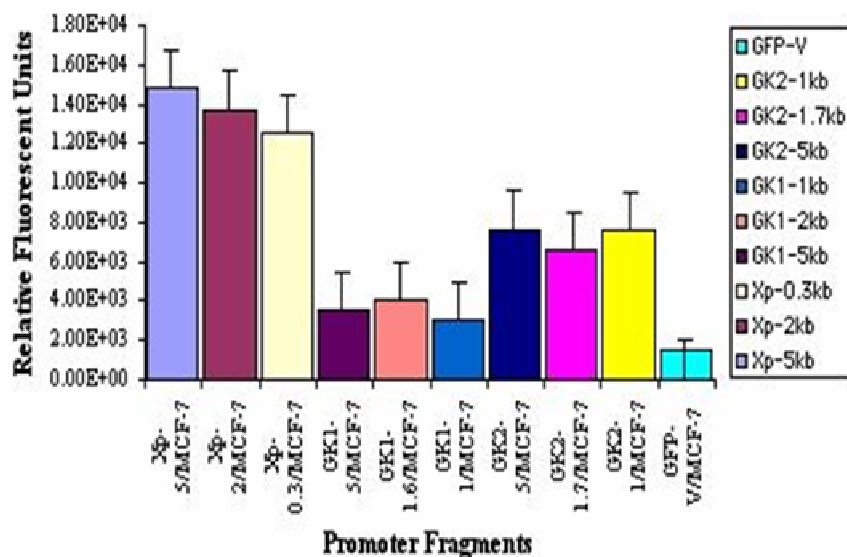


Figure 3D. GK gene family promoters in MCF-7 cells.

Figure 3A-D. Comparative GFP expression from GK, GK1 and GK2 promoters in Hep G2 cells, MCF-7, GC1 and GC2 cells. The GK1 and GK2 promoters were most active in the mouse GC1 and GC2 germ cells (fig.3A and B), whereas GK-Xp was most active in Hep G2 and MCF-7 cells (fig.3C and D). The error bar is the standard deviation from the averaged fluorescence from the three transfections per construct.

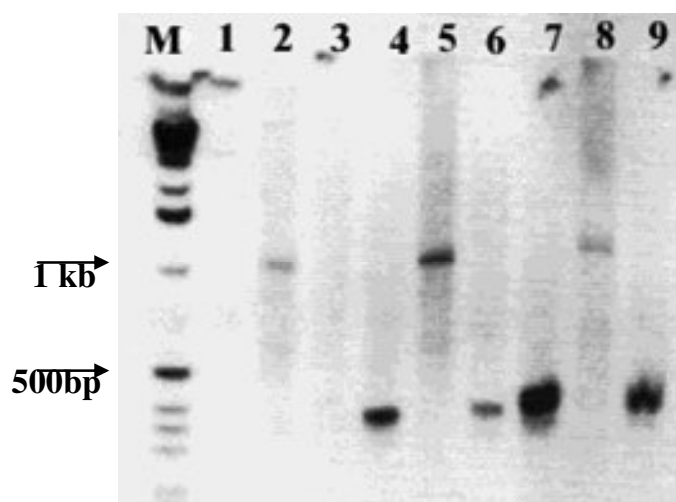


Figure 4. Glycerol kinase gene family RACE PCR. Lane M is molecular weight marker. Lanes 1 and 3 = GK Xp, as negative control. Lanes 2, 5 and 8 = positive controls. Lanes 4 and 6 = GK-1 (142). Lanes 7 and 9 = GK-2 (A13). RACE PCR analysis on 2% Agarose gel. Lane M is the molecular weight marker. Lanes 2, 5, and 8 are positive controls amplified with primers supplied by the manufacturer (Clontech™). Lanes 1 and 3 are negative control PCRs with primers derived from the non-testis GK-Xp gene. Lanes 4 and 6 are the GK-1 (142) gene RACE products. Lanes 7 and 9 are the GK-2 (A13). The negative controls in lane 3 and the positive control in lanes 2 and 8 were amplified at 68°C. The negative control in lane 1, and the positive control in lanes 5, and 8 were amplified at T_m 72°C, The GK-1, and GK-2 were amplified at T_m 72°C respectively.

functions. Most promoter studies were previously done *in vitro*, using testis whole cell nuclear extracts in run-off transcription assays. The first immortalised germ cells were made using mouse cells (Hofmann et al., 1992). The GC-1spg (GC1 cells) represented an immortalised pachytene stage primary spermatocytes. The cells were isolated from primary murine pachytene spermatocytes trans-fected with SV40 Large T antigen (Hofmann et al., 1992). The cells do not undergo differentiation in culture, or even in co-culture with immortalised Sertoli cells. Prior to their utilisation for *in situ* promoter studies, the cells were genetically characterised and were found to express *Pgk-2* and protamine transcripts (data not shown). *Pgk-2* transcription is known to occur only in post-meiotic germ cells (Feig et al., 2007), while protamine is only expressed in round spermatids stage (Hecht, 1986). To understand testis-specific gene regulation during spermatogenesis, it is essential to analyse in detail the *cis*-elements in pro-moters that are expressed at similar stages of spermatogenesis.

In the analysis of the interaction of purified TFIID with several TATA-less promoters, a conserved downstream core promoter element was found that required sequence-specific binding of TFIID to a subset of TATA-less promoters ((Burke, 1996)); this motif was called DPE, and it is a distinct 7-nucleotide element that is located at about + 30 (typically + 27 to + 34) relative to the transcription start site (consensus, $^A/G^A/T^A/CGTG$). The DPE motif in the *GK-1*, sequence GGACGTG, is found at + 27 position, whereas the one in *GK-2*, se-

quence AGAGCTG, is located at + 30 position. The DPE motif is conserved from *Drosophila* to humans and has been identified specifically only in initiator element containing TATA-less promoters, although not all TATA-less promoters contain DPE, TATA containing promoters generally do not contain DPE. Human and *Drosophila* transcription factors exhibit essentially the same requirement for DPE sequence and for the Inr-DPE spacing. The motif specifically binds the TFIID but not TBP, thereby recruiting the RNA Pol II to the transcription start site via interaction with the Inr (Jayakumar et al., 2007).

Perhaps most interesting are the motifs *GK-1* and *GK-2* have in common with other testis-specific genes, or that are known to play significant role in the regulation of transcription in other testis specific metabolic enzyme genes. This includes the Ets motif, which has been found in a number of testis-specific metabolic enzyme genes. Analysis of the mouse *Pdha-2* regulatory region indicates the presence of an Ets element, at -130 to -121 positions proximal to the transcription start site (Iannello et al., 2000) that interacts with an Ets-related factor (TAP-1). A transcription factor, TIN-1, which acts as a repressor in somatic tissues (Goto et al., 1991); (Goto et al., 1993) has previously been reported bound to a sequence similar to the Ets binding site. Furthermore, a unique Ets-1 transcription factor, ER71, that binds to this motif and expressed only in testicular tissue has been identified (Brown and Mcknight, 1992) and has been suggested as candidate molecule for TAP-1. Furthermore, the Ets-1 transcription factor has been shown to act synergistically with Sp1 (Gegonne et al., 1993), suggesting that cooperative interaction between these factors is important in transcriptional activation, perhaps with TIN-1 repressing *GK-1* and *GK-2* in somatic tissues while TAP-1 activates them in germ cells. Remarkably, an Sp1 binding site is located just downstream of the Ets 1 motif in *GK-1* and *GK-2*, as well as in both mouse *Pdha-1* and *Pgk-2*, promoters.

In an *in vitro* test, using testis nuclear extracts, the removal of the Ets motif caused a loss of activity in the mouse *Pgk-2* promoter (Goto et al., 1993), and increased activity in the human *PDHA-2* promoter (Datta et al., 1999). This suggests that Ets 1 function probably depended on the context of other transcriptional factors with which it interacts in a particular promoter. Another notable finding in the *GK-1* and *GK-2* promoters is the MEP-2 binding element, this motif has been identified in most testis specific metabolic enzyme gene promoters, suggesting regulation by a common transcription factor(s). Analysis of the MEP-2 binding domain in a number of testis-specific gene promoters revealed its presence in *PGK-2* (Gebara and McCarrey, 1992), *Pdha-2* (Iannello et al., 1995), *TH2B* (Lim and Chae, 1992), and *Zfp-35* (de Luis et al., 1999). In all the promoters analysed, the sequence was found in core promoter regions. There was a wide variation from the MEP-2 consensus sequence among these promoters. The motif

is bound by E1/E2 transcription factors and factor(s) found exclusively in testis nuclear extracts, suggesting a tissue specific regulatory function. What these factors are and how they modulate transcription is yet to be determined.

In addition to the metabolic enzyme genes examined, protamine is perhaps the most extensively studied amongst testis-specific genes. *Prm-1* initiates transcription in round spermatids, as opposed to the pachytene cells for the metabolic enzymes (Hecht, 1986). However, comparison of the regulatory regions of *PGK-2* and *Prm-1* revealed sequence elements that may be involved in tissue-specific expression. Both *Prm-1* and human *PGK-2* 5' regions contain two copies each of the sequence GGGTGGGG, present in opposite orientation in the two genes. This sequence is also found in the upstream region of the *Prm-2* gene, in the same orientation as in *PGK-2* (McCarrey, 1987). Interestingly, analysis of the *GK-1* and *GK-2* promoter sequences revealed that *GK-1* does not contain the motif, while *GK-2* does. The *GK-2* promoter contains two copies of the motif, as found in *Prm-1* and *PGK-2*, however, in the same orientation as in *PGK-2*. The motifs are located at - 87 to -94 and - 1832 to - 1839 positions in the *GK-2* sequence. It should be noted that this motif has not been found in the available mouse *Pgk-2* gene.

The cyclic-AMP response element, CRE, is perhaps the most studied transcriptional motif amongst the testis germ cell-specific transcriptional elements. The CRE element is an enhancer that mediates both basal and cAMP-induced transcriptional activation of variety of genes in numerous cell types (Fimia et al., 2000). CRE, however, is also the binding site for the testis-specific transcriptional activator CREM tau which regulates the expression of a number of testis-specific post-meiotic genes (Kistler et al., 1994). Nevertheless, CRE does not feature prominently as a regulator of the testis-specific isoforms of the metabolic enzymes, as most of the promoters do not contain the motif and its role in those that did has not been specifically determined. The CRE motif in the human *LDHC* promoter, for instance, binds to testis-specific and ubiquitous factors in mobility shift assays (Bavner et al., 2005). In *GK-1* and *GK-2*, there are CRE-like motifs at - 29 in *GK-1* and at - 28 and - 12 in *GK-2*. It is not known if these motifs represent genuine binding sites.

Conclusion

Finally, it is clear that various elements modulate tissue specific expression of GK gene family. Some of the elements that may be involved in this modulation have been identified, particularly those GK shares with other testis specific metabolic enzyme genes. These elements, such Ets and MEP-2 probably play a major role in regulating these genes, based on their presence in almost all the testis specific gene promoters. Further-

more, the MEP-2 motif is found in both testis-specific *GK* gene promoters, including other testis-specific metabolic enzyme genes. This suggests that MEP-2, like the CREM proteins, may similarly regulate these genes. However, this determination will have to await further studies, such as *in vitro* competition assays. Gel mobility shift assays will need to be done to determine what factors in the cells are responsible for modulating the tissue-specific expressions of these promoters, as well as identifying additional motifs that may be important in such modulation. Currently work is underway to generate transgenic mice with some of the constructs discussed in these studies. This will help to delineate more precisely the function of these promoters *in vivo*.

REFERENCE

- Bavner A, Matthews J, Sanyal S, Gustafsson JA, Treuter E (2005). EID3 is a novel EID family member and an inhibitor of CBP-dependent co-activation. *Nucleic Acid Res.* 24; 33(11): 3561-3569
- Brown TA, Mcknight SL (1992). Specificities of protein-protein and protein-DNA interaction of GABP alpha and two newly defined ets-related proteins. *Gene Dev.* (12B): 2502-2512.
- Bucher P (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* 20: 212(4): 563-578.
- Churikov D, Siino J, Svetlova M, Zhang K, Geneitis AM, Bradbury EZ (2004). A Novel human testis-specific histone H2B encoded by the interrupted gene on the X chromosome. *Genomics*; 84(4): 745-756.
- Datta, U, Wexler, ID Kerr, DS, Raz I, Patel MS (1999). Characterization of the regulatory region of the human testis-specific form of the pyruvate dehydrogenase a-subunit (PDHA-2) gene. *Biochem. Biophys. Acta* 1447: 236-243.
- de Luis O, Lopez-fernandez LA, del Mazo J (1999). Gene containing a zinc-finger domain, is up-regulated during the haploid stages of spermatogenesis. *Exp. Cell Res.* 5; 249(2): 320-326.
- Feig C, Kirchhoff C, Ivell R, Naether O, Schulze W, Spiess AN (2007). A new paradigm for profiling testicular gene expression during normal and disturbed human spermatogenesis. *Mol. Hum. Reprod.*, 13(1): 33-43.
- Fimia GM, De Cesare D, Sassone-Corsi P (2000). A family of LIM-only transcriptional coactivators: tissue-specific expression and selective activation of CREB and REM. *Mol. Cell Biol.*, 20(22): 8613-8622.
- Gebara MM, McCarrey JR (1992). Protein-DNA Interactions Associated with the Onset of Testis-Specific Expression of the Mammalian Pkg-2 Gene. *Mol. Cell. Biol.* 12: 1422-1431.
- Gegonne A, Bosselut R, Bailly RA, Ghysdael J (1993). Synergistic activation of the HTLV1 LTR Ets-responsive region by transcription factors Ets 1 and Sp1. *EMBO* 12: 1169-1178.
- Goto M, Masamune Y, Nakanishi Y (1993). A factor stimulating transcription of the testis-specific Pkg-2 gene recognizes a sequence similar to the binding site for a transcription inhibitor of the somatic-type Pkg-1 gene. *Nucleic Acids Res.* 21: 209-214.
- Goto M, Tamura TA, Mikoshiba K, Masamune Y, Nakanishi Y (1991). Transcription inhibition of the somatic-type phosphoglycerate kinase 1 gene *in vitro* by a testis-specific factor that recognizes a sequence similar to the binding site for Ets oncoprotein. *Nucleic Acids Res.* 19: 3959-3963.
- Guo W, Worley K, Adams V, Mason J, Sylvester-Jackson D, Zhang YH, Twbin JA, Fogt DD, Madu S, Wheeler DA, McCabe ERB (1993). Genomic Scanning for expressed sequences in Xp21 identifies the glycerol kinase gene. *Nature Genet.* 4: 367-372.
- Hofmann MC, Narisawa S, Hess RA, Millán JL (1992). Immortalization of germ cells and somatic testicular cells using the SV40 large T antigen. *Exp. Cell Res.* 201: 2417-2435.
- Iannello RC, Young JC, Sumarsono S, Tymms MJ, Kola I (1994). Mouse testis Pdha-2 promoter upstream sequences confer tissue- and temporal-specific activity in transgenic mice. *Reprod. Fert. Dev.* 6: 599-604.
- Jayakumar PC, Shouche YS, Patole MS (2007). Functional analysis of *Drosophila melanogaster* hexokinase Hex-A locus: multiple Initiator-like elements enhance DPE containing promoter activity. *Insect Mol. Biol.*, 16(1): 3-13.
- Kistler M, Sassone-Corsi P, Kistler SW (1994). Identification of functional cAMP response element in the 5' flanking region of the gene for transition protein 1 (TP1), a basic chromosomal protein of mammalian spermatids. *Biol. Reprod.* 51: 1322-1329.
- Lim K, Chae CB (1992). Presence of a repressor protein for testis-specific H2B (TH2B) histone gene in early stages of spermatogenesis. *J. Biol. Chem.* 267: 15271-15273.
- McCabe E (1983). Human glycerol kinase deficiency: An inborn error of compartmental metabolism. *Biochem. Med.* 30: 215-230.
- McCabe ERB (1989). Disorders of glycerol metabolism. *The Metabolic Basis of Inherited Disease. C. R. e. a. Scriver.* New York, McGraw-Hill, pp. 945-961.
- McCarrey JR (1987). Nucleotide sequence of the promoter region of a tissue-specific human retroposon: comparison with its housekeeping progenitor. *Gene* 61: 291-298.
- McCarrey JR, Kumari M, Aivaliotis MJ, Wang Z, Zhang P, Marshall F, Vazdeberg JL (1996). Analysis of the cDNA and encoded protein of the human testis-specific PGK-2 gene. *Dev Genet.*, 19(4): 321-332.
- Pan Y, Decker WK., Huq AHM, Craigen WJ (1999). Retrotransposition of Glycerol Kinase-Related Genes from the X Chromosome to Autosomes: Functional and Evolutionary Aspects. *Genomics* 59: 282-290.
- Roeder RG (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* 21: 327-335.
- Sargent CA, Young C, Marsh S, Ferguson-Smith MA, Affara NA (1994). The glycerol kinase gene family: Structure of the Xp gene, and related intronless retroposons. *Hum. Mol. Gen.*, 3: 1317-1324.
- Singh U, Rogers BJ, Mann JB, Petri W (1997). Transcription initiation is controlled by three core promoter elements in the *hg15* gene of the protozoan parasite *Entamoeba histolytica*. *Proc. Natl. Acad. Sci. USA* 94: 8812-8817.
- Smale ST, Baltimore D (1989). The "initiator" as a transcription control element. *Cell* 57: 103-113.
- Walker AP, Muscatelli F, Monaco AP (1993). Isolation of the human Xp21 glycerol kinase gene by positional cloning. *Hum. Mol. Genet.* 2: 107-114.
- Wilson JE (1978). Ambiquitous enzymes: Variation in intracellular distribution as a regulatory mechanism. *Trends Biochem. Sci.*, 3: 124-124.
- Ye J, Young HA, Zhang X, Castranova V, Vallyathan V, Shi X (1999). Regulation of a CellType-specific Silencer in the Human Interleukin-3 Gene Promoter by the Transcription Factor YY1 and an AP2 Sequence-recognizing Factor. *J. Biol. Chem.* 274: 26661-26667.