The HNF-3 Gene Family of Transcription Factors in Mice: Gene Structure, cDNA Sequence, and mRNA Distribution

KLAUS H. KAESTNER, HOLGER HIEMISCH, BRUNO LUCKOW,[1] AND GÜNTHER SCHÜTZ[2]

Division of Molecular Biology of the Cell I, German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

The rat HNF-3 (hepatocyte nuclear factor 3) gene family encodes three transcription factors known to be important in the regulation of gene expression in liver and lung. We have cloned and characterized the mouse genes and cDNAs for HNF-3α, β, and γ and analyzed their expression patterns in various adult tissues and mouse embryonic stages. The HNF-3 proteins are highly conserved between mouse and rat, with the exception of the amino terminus of HNF-3γ, which in mouse is more similar to those of HNF-3α and β than to the amino termini of the rat HNF-3γ protein. The mouse HNF-3 genes are small and contain only two or three (HNF-3β) exons with conserved intron–exon boundaries. The proximal promoter of the mouse HNF-3β gene is remarkably similar to that of the previously cloned rat HNF-3β gene, but is different from the promoters of the HNF-3α and γ genes. The mRNA distribution of the mouse HNF-3 genes was analyzed by quantitative RNase protection with gene-specific probes. While HNF-3α and β are restricted mainly to endoderm-derived tissues (lung, liver, stomach, and small intestine), HNF-3γ is more extensively expressed, being present additionally in ovary, testis, heart, and adipose tissue, but missing from lung. Transcripts for HNF-3β and α are detected most abundantly in midgestation embryos (Day 9.5), while HNF-3γ expression peaks around Day 15.5 of gestation. © 1994 Academic Press, Inc.

INTRODUCTION

Much progress has been made toward understanding liver-specific gene expression through biochemical analysis of the promoters and enhancers of genes such as those encoding albumin or transthyretin and the subse-

quent isolation of the transcription factors important in their regulation (for review see Lai and Darnell, 1991; Sladek and Darnell, 1992; De Simone and Cortese, 1991).

The major liver-enriched transcription factors identified so far include HNF-1, a POU-homeodomain protein; C/EBPα and β, which are bZip proteins; HNF-4, a member of the thyroid–steroid hormone receptor superfamily; and HNF-3, a representative of the HNF-3/forkhead class of DNA-binding proteins (Lai and Darnell, 1991; Sladek and Darnell, 1992; De Simone and Cortese, 1991, and references therein). The HNF-3 proteins were first identified by their ability to bind to important promoter elements in the α_1-antitrypsin and transthyretin genes (Costa *et al.*, 1989). Other target sites for HNF-3 have been described in the α-fetoprotein, albumin, tyrosine aminotransferase, phosphoenolpyruvate kinase, transferrin, and HNF-1α and HNF-3β genes (Costa *et al.*, 1989; Herbst *et al.* 1991; Nitsch *et al.*, 1993; Ip *et al.*, 1990; Auge-Gouillou *et al.*, 1993; Kuo *et al.*, 1992; Pani *et al.*, 1992b) as well as the lung-specific gene CC10 (Sawaya *et al.*, 1993). From comparison of the observed HNF-3 binding sites, a consensus element, TATTGA$^C/_T$TT$^A/_T$G, has been derived (Costa *et al.*, 1989), although deviations from this consensus sequence have been identified (Nitsch *et al.*, 1993; Pani *et al.*, 1992b). The HNF-3 DNA-binding activities were purified from rat liver extracts and the rat cDNAs subsequently cloned and sequenced (Lai *et al.*, 1990, 1991). The three binding activities observed in gel retardation assays were found to correspond to three proteins encoded by independent genes, which were termed HNF-3α, β, and γ (Lai *et al.*, 1991). Deletion analysis of HNF-3α demonstrated that the region between amino acids 124 and 288 is essential for DNA binding (Lai *et al.*, 1990). Subsequent sequence comparison revealed striking similarities between HNF-3 and the *Drosophila melanogaster* gene forkhead within the DNA binding domain (Lai *et al.*, 1991; Weigel and Jäckle, 1990), for which 100 of 110 amino acids are identical between HNF-3β and forkhead. The *Drosophila* gene forkhead has been described as a region-specific homeotic gene and is required for the proper formation of the terminal structures of the *Drosophila* embryo

```
A    1  CGCCGCGCCGCGCCGCCGCCGCCGCGCACGCCGCGCCCCGCAGCGCCGGGCTTCCTCCTCGCCCGGGTGGCGCTGGGCCCTCGAGCGCTCCGGTGACCGCAGCGGCTCCGCGCCCCTCCC  120

   121  CCGCCCCGAGCAGCGCACCCGCCCGTCGCTCCGCACAGGGTTGGATGGTTGTGTCGGCCGGGCTGGCTCCAGGATGTTAGGGACTGTGAAGATGGAAGGGCATGAGAGCAACGACTGGAA  240
     1                                                                 M  L  G  T  V  K  M  E  G  H  E  S  N  D  W  N     16

   241  CAGCTACTACGCGGACACGCAGGAGGCCTACTCCTCTGTCCCTGTCAGCAACATGAACTCCGGCCTGGGCTCTATGAACTCCATGAACACCTACGTGACCATGAACACCATGACCACGAG  360
    17  S  Y  Y  A  D  T  Q  E  A  Y  S  S  V  P  V  S  N  M  N  S  G  L  G  S  M  N  S  M  N  T  Y  M  T  M  N  T  M  T  T  S   56

   361  CGGCAACATGACCCCGGCTTCCTTCAACATGTCCTACGCCAACACGGGCTTAGGGGCCGGCCTGAGTCCCGGTGCTGTGGCTGGCATGCCAGGGGCCTCTGCAGGCGCCATGAACAGCAT  480
    57  G  N  M  T  P  A  S  F  N  M  S  Y  A  N  T  G  L  G  A  G  L  S  P  G  A  V  A  G  M  P  G  A  S  A  G  A  M  N  S  M   96

   481  GACTGCGGCGGGCGTCACGGCCATGGGTACGGCGCTGAGCCCGGGAGGCATGGGCTCCATGGGCGCGCAGCCCGTCACCTCCATGAACGGCCTGGGTCCCTACGCCGCCGCCATGAACCC  600
    97  T  A  A  G  V  T  A  M  G  T  A  L  S  P  G  G  M  G  S  M  G  A  Q  P  V  T  S  M  N  G  L  G  P  Y  A  A  A  M  N  P   136

   601  GTGCATGAGTCCCATGGCGTACGCGCCGTCCAACCTGGGCCGCAGCCGCGCGGGGGGCGGCGGCGACGCCAAGACATTCAAGCGCAGCTACCCTCACGCCCAAGCCGCCTTACTCCTACAT  720
   137  C  M  S  P  M  A  Y  A  P  S  N  L  G  R  S  R  A  G  G  G  G  D  A  K̲  T̲  F̲  K̲  R̲  S̲  Y̲  P̲  H̲  A̲  K̲  P̲  P̲  Y̲  S̲  Y̲ I̲  176

   721  CTCGCTCATCACGATGGCCATCCAGCAGGCGCCCAGCAAGATGCTCACGCTGAGCGAGATCTACCAGTGGATCATGGACCTCTTCCCCTATTACCGCCAGAACCAGCAGCGCTGGCAGAA  840
   177  S̲ L̲ I̲ T̲ M̲ A̲ I̲ Q̲ Q̲ A̲ P̲ S̲ K̲ M̲ L̲ T̲ L̲ S̲ E̲ I̲ Y̲ Q̲ W̲ I̲ M̲ D̲ L̲ F̲ P̲ Y̲ Y̲ R̲ Q̲ N̲ Q̲ Q̲ R̲ W̲ Q̲ N̲  216

   841  CTCCATCCGCCACTCGCTGTCCTTCAACGATTGTTTCGTCAAGGTGGCACGATCCCCAGACAAGCCAGGCAAGGGCTCCTACTGGACGCTGCACCCGGACTCCGGCAACATGTTCGAGAA  960
   217  S̲ I̲ R̲ H̲ S̲ L̲ S̲ F̲ N̲ D̲ C̲ F̲ V̲ K̲ V̲ A̲ R̲ S̲ P̲ D̲ K̲ P̲ G̲ K̲ G̲ S̲ Y̲ W̲ T̲ L̲ H̲ P̲ D̲ S̲ G̲ N̲ M̲ F̲ E̲ N̲  256

   961  CGGCTGCTACTTGCGCCGCCAAAAGCGCTTCAAGTGTGAGAAGCAGCCGGGGGCCGGAGGTGGGAGTGGGGGCGGCGGCTCCAAAGGGGGCCCAGAAAGTCGCAAGGACCCCTCAGGCCC  1080
   257  G̲ C̲ Y̲ L̲ R̲ R̲ Q̲ K̲ R̲ F̲ K̲ C̲ E̲ K̲  Q  P  G  A  G  G  G  S  G  G  G  G  S  K  G  G  P  E  S  R  K  D  P  S  G  P   296

  1081  GGGGAACCCCAGCGCCGAGTCACCCCTTCATTGGGGTGTGCACGGAAAGGCTAGCCAGCTAGAGGGCGCGCCGGCCCCCGGGCCCGCCGCCAGCCCCCAGACTCTGGACCACAGCGGGGC  1200
   297  G  N  P  S  A  E  S  P  L  H  W  G  V  H  G  K  A  S  Q  L  E  G  A  P  A  P  G  P  A  A  S  P  Q  T  L  D  H  S  G  A   336

  1201  CACGGCCGACAGGGGGCGCTTCGGAGTTGAAGTCTCCAGCGTCTTCATCTGCGCCCCCCATAAGCTCCGGGCCAGGGGCGCTGGCATCTGTACCCCCCTCTCACCCGGCTCACGGCCTGGC  1320
   337  T  A  T  G  G  A  S  E  L  K  S  P  A  S  S  A  P  P  I  S  S  G  P  G  A  L  A  S  V  P  P  S  H  P  A  H  G  L  A   376

  1321  ACCCCACGAATCTCAGCTGCATCTGAAAGGGGATCCCCACTACTCCTTTAATCACCCCTTCTCCATCAACAACCTCATGTCCTCCTCCGAGCAACAGCACAAGCTGGACTTCAAGGCATA  1440
   377  P  H  E  S  Q  L  H  L  K  G  D  P  H  Y  S  F  N  H  P  F  S  I  N  N  L  M  S  S  S  E  Q  Q  H  K  L  D  F  K  A  Y   416

  1441  CGAGCAGGCGCTGCAGTACTCTCCTTATGGCGCTACCTTGCCCGCCAGTCTGCCCCTTGGCAGCGCCTCAGTGGCCACGAGGAGCCCCATCGAGCCCTCAGCCCTGGAGCCAGCCTACTA  1560
   417  E  Q  A  L  Q  Y  S  P  Y  G  A  T  L  P  A  S  L  P  L  G  S  A  S  V  A  T  R  S  P  I  E  P  S  A  L  E  P  A  Y  Y   456

  1561  CCAAGGTGTGTATTCCAGACCCGTGCTAAATACTTCCTAGCTCCCAGAACTGAGGGTTTTGTCTGCATGGCCAACCTGGTAGCAGGAGAGAAAAACCAACAGCAAACAAACAAAACCACA  1680
   457  Q  G  V  Y  S  R  P  V  L  N  T  S  *                                                                                469

  1681  CACACACACACAGACACACACACACCAAACCAACCACAT**AATAAAA**TTCCAAATAGTTTTATTTTTCACGCACAAAAAAAAAA  1766
```

```
B    1  CTGACGACCAGGGCGGCCAGACCACGCGAGTCCTACGCGCCTCCTGAGGCCGCCCCGGGACTTAACTGTAACGGGGAGGGGCCTCCGGAGCAGCGGCCAGCGAGTTAAAGTATGCTGGGA  120
     1                                                                                                  M  L  G   3

   121  GCCGTGAAGATGGAAGGGCTCGAGCCCATCCGACTGGAGCAGCTACTACGCGGAGCCCGAGGGCTACTCTTCCGTGAGCAACATGAACGCCGGCCTGGGGATGAATGGCATGAACACATAC  240
     4  A  V  K  M  E  G  L  E  P  S  D  W  S  S  Y  Y  A  E  P  E  G  Y  S  S  V  S  N  M  N  A  G  L  G  M  N  G  M  N  T  Y   43

   241  ATGAGCATGTCCGCGGCTGCCATGGGCGGCCGGTTCCGGCAACATGAGCGCGGGCTCCATGAACATGTCATCCTATGTGGGCGCTGGAATGAGCCCGTCGCTAGCTGGCATGTCCCCGGGC  360
    44  M  S  M  S  A  A  A  M  G  G  G  S  G  N  M  S  A  G  S  M  N  M  S  S  Y  V  G  A  G  M  S  P  S  L  A  G  M  S  P  G   83

   361  GCCGGCGCCATGGCGGGCATGAGCGGCTCAGCCGGGGCGGCCGGCGTGGCGGGCATGGGACCTCACCTGAGTCCGAGTCTGAGCCCGCTCGGGGGGCAGGCGGCCGGGGGCCATGGGTGGC  480
    84  A  G  A  M  A  G  M  S  G  S  A  G  A  A  G  V  A  G  M  G  P  H  L  S  P  S  L  S  P  L  G  G  Q  A  A  G  A  M  G  G   123

   481  CTTGCCCCCTACGCCAACATGAACTCGATGAGCCCCATGTACGGGCAGGCCGGCCTGAGCCGCGCTCGGGACCCCAAGACATACCGACGCAGCTACACACACGCCCAAACCTCCCTACTCG  600
   124  L  A  P  Y  A  N  M  N  S  M  S  P  M  Y  G  Q  A  G  L  S  R  A  R  D  P  K̲  T̲  Y̲  R̲  R̲  S̲  Y̲  T̲  H̲  A̲  K̲  P̲ P̲ Y̲ S̲  163

   601  TACATCTCGCTCATCACCATGGCCATCCAGCAGAGCCCCAACAAGATGCTGACGCTGAGCGAGATCTATCAGTGGATCATGGACCTCTTCCCCTTTCTACCGGCAGAACCAGCAGCGCTGG  720
   164  Y̲ I̲ S̲ L̲ I̲ T̲ M̲ A̲ I̲ Q̲ Q̲ S̲ P̲ N̲ K̲ M̲ L̲ T̲ L̲ S̲ E̲ I̲ Y̲ Q̲ W̲ I̲ M̲ D̲ L̲ F̲ P̲ F̲ Y̲ R̲ Q̲ N̲ Q̲ Q̲ R̲ W̲  203

   721  CAGAACTCCATCCGCCACTCTCTCTCCTTCAACGACTGCTTTCTCAAGGTGCCCCGCTCGCCAGACAAGCCTGGCAAGGGCTCCTTCTGGACCCTGCACCCAGACTCGGGCAACATGTTC  840
   204  Q̲ N̲ S̲ I̲ R̲ H̲ S̲ L̲ S̲ F̲ N̲ D̲ C̲ F̲ L̲ K̲ V̲ P̲ R̲ S̲ P̲ D̲ K̲ P̲ G̲ K̲ G̲ S̲ F̲ W̲ T̲ L̲ H̲ P̲ D̲ S̲ G̲ N̲ M̲ F̲  243

   841  GAGAACGGCTGCTACCTGCGCCGCCAGAAGCGCTTCAAGTGTGAGAAGCAACTGGCACTGAAGGAAGCCGCGGGTGCGGCCAGTAGCGGAGGCAAGAAGACCGCTCCTGGGTCCCAGGCC  960
   244  E̲ N̲ G̲ C̲ Y̲ L̲ R̲ R̲ Q̲ K̲ R̲ F̲ K̲ C̲ E̲ K̲  Q  L  A  L  K  E  A  A  G  A  A  S  S  G  G  K  K  T  A  P  G  S  Q  A   283

   961  TCTCAGGCTCAGCTCGGGGAGGCCGCGGGCTCGGCCTCCGAGACTCCGGCGGGCACCGAGTCCCCCATTCCAGCGCTTCTCCGTGTCAGGAGCACAAGCGAGGTGGCCTAAGCGAGCTA  1080
   284  S  Q  A  Q  L  G  E  A  A  G  S  A  S  E  T  P  A  G  T  E  S  P  H  S  S  A  S  P  C  Q  E  H  K  R  G  G  L  S  E  L   323

  1081  AAGGGAGCACCTGCCTCTGCGCTGAGTCCTCCCGAGCCGGCGGCCTCGCCTGGGCAGCAGCAGCAGGCTGCAGCCCACCTGCTGGGCCCCACCTCACCACCCAGGCCTGCCACCAGAGGCC  1200
   324  K  G  A  P  A  S  A  L  S  P  P  E  P  A  P  S  P  G  Q  Q  Q  Q  A  A  A  H  L  L  G  P  P  H  H  P  G  L  P  P  E  A   363

  1201  CACCTGAAGCCCGAGCACCATTACGCCTTCAACCACCCCTTCTCTATCAACAACCTCATGTCGTCCGAGCAGCAACATCACCACAGCCACCACCACCATCAGCCCCACAAAATGGACCTC  1320
   364  H  L  K  P  E  H  H  Y  A  F  N  H  P  F  S  I  N  N  L  M  S  S  E  Q  Q  H  H  H  S  H  H  H  H  Q  P  H  K  M  D  L   403

  1321  AAGGCCTACGAACAGGTCATGCACTACCCAGGGGGCTATGGTTCCCCCATGCCAGGCAGCTTGGCCATGGGCCCAGTCACGAACAAAGCGGGCCTGGATGCCTCGCCCCTGGCTGCAGAC  1440
   404  K  A  Y  E  Q  V  M  H  Y  P  G  G  Y  G  S  P  M  P  G  S  L  A  M  G  P  V  T  N  K  A  G  L  D  A  S  P  L  A  A  D   443

  1441  ACTTCCTACTACCAAGGAGTGTACTCCAGGCCTATTATGAACTCATCCTAAGAAGATGGCTTTCAGGCCTGCTAGCTCTGGTCACTGGGGACAAGGGAAATGAGAGGCTGAGTGGAGAC  1560
   444  T  S  Y  Y  Q  G  V  Y  S  R  P  I  M  N  S  S  *                                                                     460

  1561  TTTGGGAGAGCTTTGAGGAAAAGTAGCCACCACACTTCAGGCCTCAAGGGAGCAGTCTCACCTGTCTGTGTCCCTAAATAGATGGGCCACAGTGATCTGTCATTCTAAATAGGGAAGGGA  1680

  1681  ATGGAAATATATATGTATACATATAAACTTGTTTTAAAGGAGCCTTTGGTCTCCTCTATGTAGACTACTGCTTCTCAAGACATCTGCAGAGTTTGATTTTTGTTGTTGTTCTCTATTGCT  1800

  1801  GTTGTTGCAGAAAAGTCTGACTTTAAAAACAAACAAACAAACAAAAAACTTTTGTGAGTGACTTGGTGTAAAACCATGTAGTTTTAACAGAAAACCAGAGGGTTGTACTGATGTTGAAAA  1920

  1921  GAGGAAAGAAAAATAATGTAAGAGTCTGGTGTACCGGACCAGGAGAAAGGAGAAAAACACATCCCATTCTGGACATGGTGAAATCCAGGTCTCGGGTCTGATTTAATTTATGGTTTCTGC  2040

  2041  GTGCTTTATTTATGGCTTATAAATGTGTGTTCTGGCTAGAAT  2082
```

```
C   1   GCGGGACTCCCGGGCTGTGTGCCTCAGGTCGGAACTCGGGGCTAGTGCCTGTAGAGAGACCGAAGCACTCGGTTCCCCCAGGGGGCCTCAGCCTGGGTGTGTGGGGGCGCAGGCCGGGGA   120
    1                                                                                                                           M   1

  121   TGCTGGGCTCAGTGAAGATGGAGGCTCATGACCTGGCCGAGTGGAGCTACTACCCGGAGGCGGGCGAGGTGTATTCTCCAGTGAATCCTGTGCCCACCATGGCCCCTCTCAACTCCTACA   240
    2     L  G  S  V  K  M  E  A  H  D  L  A  E  W  S  Y  Y  P  E  A  G  E  V  Y  S  P  V  N  P  V  P  T  M  A  P  L  N  S  Y  M   41

  241   TGACCTTGAACCCACTCAGCTCTCCCTACCCTCCCGGAGGGGCTTCAGGCCTCCCCACTGCCTACAGGACCCCTGGCACCCCCAGCCCCCACTGCGCCCTTGGGGCCCACCTTCCCAAGCT   360
   42     T  L  N  P  L  S  S  P  Y  P  P  G  G  L  Q  A  S  P  L  P  T  G  P  L  A  P  P  A  P  T  A  P  L  G  P  T  F  P  S  L   81

  361   TGGGCACTGGTGGCAGCACCGGAGGCAGTGCTTCCGGGTATGTAGCCCCAGGGCCCGGGCTTGTACATGGAAAAGAGATGGCAAAGGGGTACCGGCGGCCACTGGCCCACGCCAAACCAC   480
   82     G  T  G  G  S  T  G  G  S  A  S  G  Y  V  A  P  G  P  G  L  V  H  G  K  E  M  A  K_G_Y_R_R_P_L_A_H_A_K_P_P   121

  481   CATATTCCTACATCTCTCTCATAACCATGGCTATTCAGCAGGCTCCAGGCAAGATGCTGACCCTGAGTGAAATCTACCAATGGATCATGGACCTCTTCCCGTACTACCGGGAGAACCAGC   600
  122   _Y_S_Y_I_S_L_I_T_M_A_I_Q_Q_A_P_G_K_M_L_T_L_S_E_I_Y_Q_W_I_M_D_L_F_P_Y_Y_R_E_N_Q_Q   161

  601   AACGTTGGCAGAACTCCATCCGGCATTCGCTGTCCTTCAATGACTGCTTCGTCAAGGTGGCACGCTCCCCAGACAAGCCAGGCAAAGGCTCCTACTGGGCCTTGCATCCCAGCTCTGGGA   720
  162   _R_W_Q_N_S_I_R_H_S_L_S_F_N_D_C_F_V_K_V_A_R_S_P_D_K_P_G_K_G_S_Y_W_A_L_H_P_S_S_G_N   201

  721   ACATGTTTGAGAACGGATGCTATCTCCGCCGGCAGAAGCGCTTCAAGCTGGAGGAGAAGGCAAAGAAAGGAAACAGCGCCACATCGGCCAGCAGGAATGGTACTGCGGGGTCAGCCACCT   840
  202   _M_F_E_N_G_C_Y_L_R_R_Q_K_R_F_K_L_E_E_ K  A  K  K  G  N  S  A  T  S  A  S  R  N  G  T  A  G  S  A  T  S   241

  841   CTGCCACCACTACAGCTGCCACTGCAGTCACCTCCCCGGCTCAGCCCCAGCCTACGCCATCTGAGCCCGAGGCCCAGAGTGGGGATGATGTGGGGGGTCTGGACTGCGCCTCACCTCCTT   960
  242     A  T  T  T  A  A  T  A  V  T  S  P  A  Q  P  Q  P  T  P  S  E  P  E  A  Q  S  G  D  D  V  G  G  L  D  C  A  S  P  P  S   281

  961   CGTCCACACCTTATTTCAGCGGCCTGGAGCTCCCGGGGGAACTAAAGTTGGATGCGCCCTATAACTTCAACCACCCTTTCTCTATCAACAACCTGATGTCAGAACAGACATCGACACCTT   1080
  282     S  T  P  Y  F  S  G  L  E  L  P  G  E  L  K  L  D  A  P  Y  N  F  N  H  P  F  S  I  N  N  L  M  S  E  Q  T  S  T  P  S   321

 1081   CCAAACTGGATGTGGGGTTTGGGGGCTACGGGGGCTGAGAGTGGGGAGCCTGGAGTCTACTACCAGAGCCTCTATTCCCGCTCTCTGCTTAATGCATCCTAGCAGCGCAATTGGGAACGCC   1200
  322     K  L  D  V  G  F  G  G  Y  G  A  E  S  G  E  P  G  V  Y  Y  Q  S  L  Y  S  R  S  L  L  N  A  S  *   354

 1201   ATGATGGGCGTGGGCTGCAACGTTCTTGGGCTCTGATCTTTCTGGTTACACTTTGCTTGTCCCATTAATTAACATCTTATTTGGTCTATTACTGTGATATGACCCATTAGCTACTGTGGT   1320

 1321   AACTGCCATGGACTCTTTGGTAGGCCTAGGGTTGGGGTATTAGGAAGGCAGATGCGTTTGGAAGTGCTGCGAAGGTGGTCATGTTGGACATATTGTGAAGGCAGTTAGACTGGTGTACTA   1440

 1441   TGAAAGCTAGCCATATTAAGTGAAGCCATTGGGTGATTGATCCACTGGGTGCCTGATGGTCGTGATGTTGGATGACACATGTCTGGTCCTTTGGATGATGTGTTGGCAATCTTGATTGAC   1560

 1561   CTGTGTTGGACATCTTGATTGACCTTTTGAGTATGTGACAGAACACATCTTCTTTGGCTCATTTTATCCTGGGATCGCCTCTTTTTTTTCCTCTTCTTTTCTTTTTCTTTTTTTCTTTT   1680

 1681   CCTTTTTTCTTTCTTTTTTCTTTTTTGGCAGACTTCTTGGTTCAGCAGATGCCAAATTGGCCACCATATCACATGGTGTCTTTTTGACATTCTGGATGCATGGAAGGTCACTGTATTGG   1800

 1801   CAAGGTGACATCTCAGCATGCTGCTATGCACCAAGATAGATGGTTACCACAGGCCTGCCATCTCCTTGGTGGAGGTTGGGTGAGAGGAAGAGGTGAGCAGACCTATCGAGTTTTTCTCTG   1920

 1921   AAGCCCATCCCCACCCTGTCTGTGAGAAAGGGCTAGTGTGGGTGTCGGGAGTTCCTACTGAGGTCAAGTTCTTGTCTGGGGCTTGGGAATACTGCCTGTGTTTGGCCATTAAAAGGCAC   2040

 2041   CATCTCCATAAAAAAAAAAAAAAAAAA   2067
```

FIG. 1. Nucleotide and translated amino acid sequences of the mouse HNF-3α, β, and γ cDNAs. Shown is the sequence of the sense strand of composite cDNAs and the corresponding amino acid sequence of the longest open reading frame for HNF-3α (A), HNF-3β (B), and HNF-3γ (C). The boldface sequences in A and C indicate potential polyadenylation signals. The sequences of the 110-amino-acid forkhead DNA binding domains are underlined.

(Jürgens and Weigel, 1988; Weigel et al., 1989). The forkhead gene is expressed in ectodermal as well as endodermal portions of the gut, the yolk nuclei, the salivary glands, and certain cells of the nervous system. Subsequently, sequences closely related to the HNF-3/forkhead DNA binding domain have been found in species ranging from yeast to man (Oliver et al., 1992; Knöchel et al., 1992; Dirksen and Jamrich, 1992; Ruiz i Altaba and Jessell, 1992; Häcker et al., 1992; Tao and Lai, 1992; Li et al., 1991, 1992; Li and Vogt, 1993; Kaestner et al., 1993; Clevidence et al., 1993). In mice, six HNF-3/forkhead homologues that have at least 57% amino acid identity within the DNA binding domain with forkhead have been described (Kaestner et al., 1993).

The rat HNF-3 genes have been shown to be expressed, in addition to liver, in stomach, intestine, and lung (Lai et al., 1990, 1991), which are all tissues derived, at least in part, from embryonic endoderm. This fact, combined with the high degree of similarity to the Drosophila gene forkhead, has led to the proposal that the HNF-3 genes are important in early endoderm and liver development in addition to their role in adult liver transcription (Lai and Darnell, 1991). This notion is supported by recent in situ hybridization studies on early postimplantation and midgestation mouse embryos (Sasaki and Hogan, 1993; Monaghan et al., 1993). The HNF-3 family members were found to be sequentially transcribed in the developing definite endoderm, HNF-3β being the first gene to be activated, followed by HNF-3α and finally γ. Interestingly, HNF-3β and α mRNAs were also found in cells of the notochord and ventral neural epithelium, suggesting additional functions for these genes in mesoderm and neural axis formation (Monaghan et al., 1993). Recent experiments suggest that the very early expression of the HNF-3 proteins in the liver primordium is responsible for the reorganization of the chromatin structure, as the HNF-3 proteins are involved in the precise positioning of nucleosomes over the albumin enhancer only in those tissues where the enhancer is active (McPherson et al., 1993). As a prerequisite to further our understanding of the importance of the HNF-3 gene family in mouse development, through promoter analysis in transgenic mice and generation of null alleles via homologous recombination in embryonic stem cells, we have cloned and characterized cDNAs and genomic fragments encoding the mouse HNF-3 genes. In addition, we have quantitatively analyzed their expression pattern in adult mouse tissues and whole mouse embryos from midgestation to term.
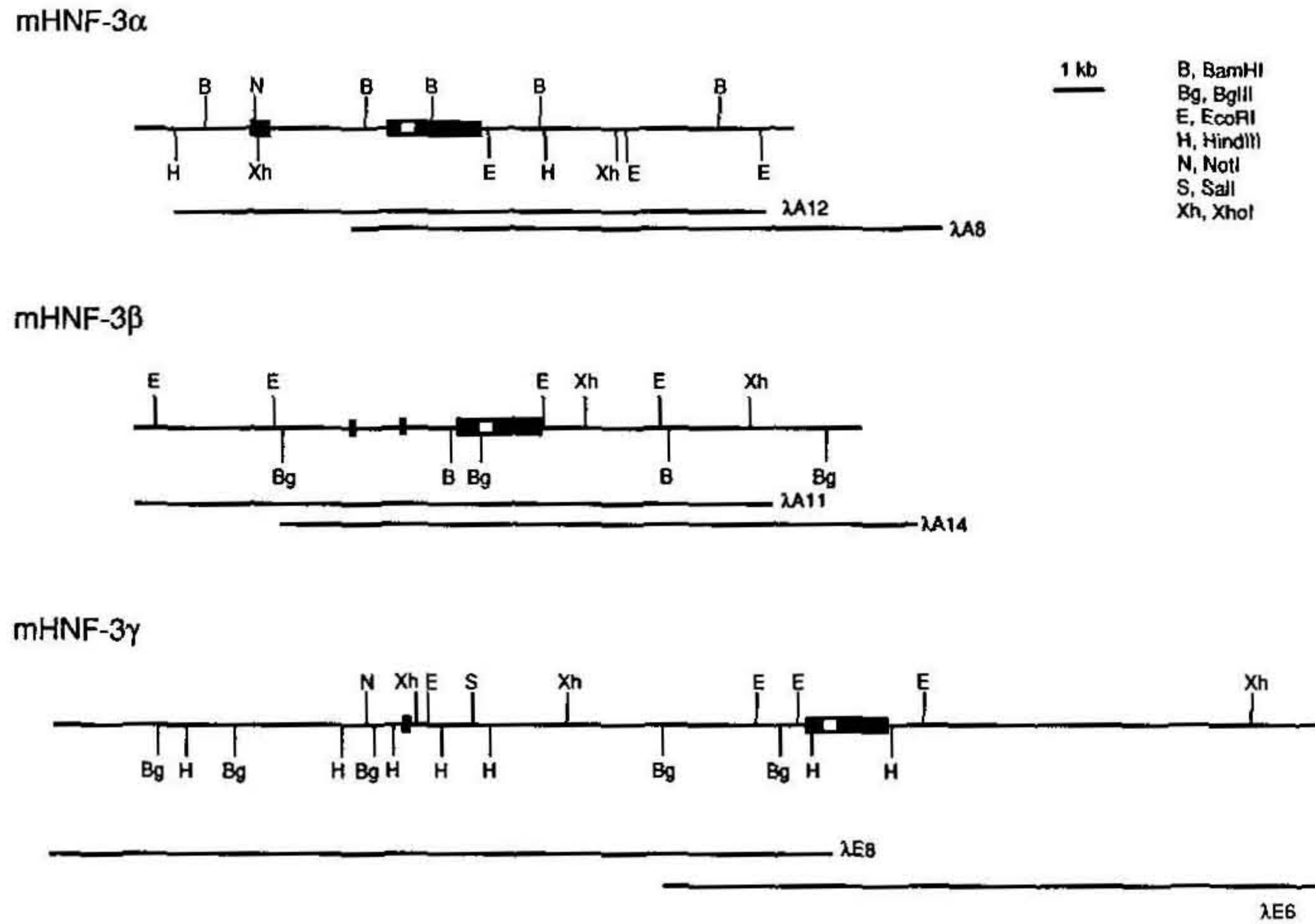
mHNF-3α

mHNF-3β

mHNF-3γ

FIG. 2. Restriction map of the mouse HNF-3α, β, and γ genes. The restriction map of each gene together with the extent of the lambda phage clones from which it was derived is shown. The exons of the three genes are shown as black boxes, and the forkhead DNA binding domains are indicated as white boxes.

MATERIALS AND METHODS

Library construction and cDNA and genomic cloning. A 300-nt probe corresponding to the forkhead domain of the mouse HNF-3α gene originally obtained by PCR (Kaestner *et al.*, 1993) was used to screen two mouse liver cDNA libraries (Ruppert *et al.*, 1990; B. Luckow *et al.*, 1994) using high-stringency hybridization and washing conditions (Church and Gilbert, 1984). Fourteen hybridizing lambda

phages were purified and the cDNAs were subcloned into Bluescript (Stratagene) and classified by dideoxy sequencing (Sanger *et al.*, 1977). The complete coding as well as part of the untranslated regions of the mouse HNF-3α, β, and γ cDNAs were sequenced after generating nested deletions of the cDNAs with exonuclease III (Henikoff, 1984).

Genomic clones containing the HNF-3α and β genes were obtained through screening of an amplified mouse 129/Ola library in λGEM-12

TABLE 1

Intron–Exon Boundaries of the HNF-3 Genes in Mice

HNF-3α

Exon 1   Exon 2
AACGCAGGAGgtgagaggcggg .... 2.5-kb intron .... tctccgctccagGCCTACTCCT
  T   Q E_{24} A_{25} Y S

HNF-3β

Exon 1   Exon 2
GCGAGTTAAAgtttttctaaat .... 1.5-kb intron .... ctcggcttccagTATGCTGGGA
                                                                      M_1 L G

Exon 2   Exon 3
GGAGCCCGAGgtaagctcttgc .... 0.5-kb intron .... ctctgtccgcagGGCTACTCTT
  E   P E_{23} G_{24} Y S

HNF-3γ

Exon 1   Exon 2
GGCGGGCGAGgtgtgtcctcgg .... 8-kb intron .... ctctccctctagGTGTATTCTC
  A   G E_{23} V_{24} Y S

Note. The intron–exon boundaries were obtained by comparison of cDNA and genomic sequences. The consensus dinucleotides gt and ag at the intron borders are underlined.

embryonic stem cell line E14TG2a (Hooper *et al.*, 1987), partially digested with the restriction endonuclease *Sau*3AI, size fractionated (16 to 23 kb fragments), and ligated into λDash II (Stratagene) according to Frischauf (1987). Two million primary phages of this library were screened with various portions of the mouse HNF-3γ cDNA to assemble the complete HNF-3γ contig. The genomic lambda phages were mapped, and exon-containing fragments subcloned and sequenced to obtain the intron–exon boundaries.

Primer extension analysis. Primer extension analysis was carried out as described previously (Kaestner *et al.*, 1989) using synthetic oligonucleotides and 5 µg tRNA or poly(A$^+$) RNA purified from stomach using oligo(dT) Dynabeads (Dynal). The primers used were complementary to nucleotides 64 to 85 of the HNF-3α cDNA, nucleotides 61 to 83 of the HNF-3β cDNA, and nucleotides 61 to 83 of the HNF-3γ cDNA (see Fig. 1). To map the endpoint of the primer-extended product, dideoxy sequencing (Sanger *et al.*, 1977) of genomic subclones containing the first exon of each gene was performed using the corresponding oligonucleotide as a primer.

RNA isolation and RNase protection analysis. Total RNA from a variety of mouse tissues or whole mouse embryos was isolated by centrifugation through a CsCl cushion after homogenization in guanidinium thiocyanate (Chirgwin *et al.*, 1979). The quality of the RNA preparations was controlled by ethidium bromide staining of the 18S and 28S rRNAs after electrophoretic separation of the RNA in denaturing agarose gels. RNase protection analysis was performed as described previously (Kaestner *et al.*, 1989) using [α − ^{32}P]UTP-labeled antisense RNA probes derived from Bluescript (Stratagene) subclones containing 191 bp (HNF-3α, position 1165 to 1355), 221 bp (HNF-3β, position 1169 to 1389), or 338 bp (HNF-3γ, position 993 to 1330) of the mouse HNF-3 cDNAs. The probes chosen were from regions outside the conserved domains of the HNF-3 family to avoid the possibility of cross-hybridization. The antisense probes were hybridized against 50 µg (adult tissues) or 20 µg (embryonic samples) total RNA

FIG. 3. Mapping of the transcription initiation sites of the mouse HNF-3α, β, and γ genes. Synthetic oligonucleotide primers complementary to the sense strand of the cDNAs were end-labeled with ^{32}P to a specific activity of 10^8 cpm/µg. Primer extension analysis was carried out with 5 µg of poly(A$^+$) RNA from stomach or 5 µg tRNA. Marker lanes contain the dideoxy nucleotide sequencing reactions of the first exon of each gene using the same primers. Arrows indicate the major start sites of transcription.

(Promega) kindly provided by Anton Berns (Amsterdam) using the probe and hybridization conditions as used for the cDNA library screen. An additional genomic library was constructed from murine embryonic stem cells. Genomic DNA was isolated from the murine

A

```
ACGATCCTTG CGCTGTAGCT CTGATGCCAC CACCCAGGTC CTTCTCCCGC AGCACAGCTC TTTGGGTCCA GAGCCCTGGC CTGTCCTCCA -396
                                                                                        SP1
AGGCACCGCC TATTTTTCCT TTTCTTCTTT TTTTTCCTTT TCTTTTCTTT CTCTTTTTTT CCTTTTCCAA AGGGGGTCAC ACACACACCC -306
                       CCAAT-BOX
GCCCTATTTT CCTCTTTCCC TAGCCGCGAG GCTTAAACCA ATTATACTGC TTTGTAAACA AAGTGAGGGC CAGGTTTGGG GGAGGGGATG -216
SP1                                              SP1                          SP1
GCGGGAGGGG CGCGGGGGGC GCGCAGGCGC TGGCGCGGCG GGGCGGGAGG CGCGGCGGCT GGACTGGCGG GCGGCCGCCT CACAGGTGCA -126

CCTCGGGCTT TGTAGGTGCG AGCGTCTTTG TGCGGCGGAC AAATGGGGAG AGGACGAGGA GGTGGGCACT CCGGCGACGT AAGATCCACA  -36
                                                  SP1 +1
TCAGCTCAAC TGCACTCGCT TCGCACAGGC CGCCCGCTCA CTTCCCGCGG AGGCGCTGCC GGGCGCCGGC TCCGCGGCCG CCTCCTGTCC  +56

CCGGCGCTGC CCCCTCCcgc cgcgccg
```

B

```
                                                                      HNF-3
CCTAGTCTCG GTCTTGGTAG CTAACAATAT AAATGACATA CTCTGTTGTT TTCATGTTTG TTTTGTTTGG GGCAGACAAG GTTTCTCTCT -365

GTAGCCGGAT GTCTTGAAAC TCACTCTATA GACTAGGTCA GCCTAGACTT CTCTGAGATC CTCCTGCCTC TGCCTCAAAG GCAGAAGACA -275

ACACTTCAAA TGACACTTTT TAACACCTAG AAGCTAAAGA GAAAAGAAAC TGAAATTTTC AATTTTTATA AGACAGTCCT GGTCTCTGCA -185
                                                              UF1-H3β                     HNF-3
GGCAGAGAAC ACAGATCCTC CTGAAGTCAT CCCACAAGGC CCATTATTGA TTTTTTCTCC TGCCCTACCC CCCACCTACT GCCCTGTTTG  -95
      LF-H3β                                        UF2-H3β
TTTTAGTTAC GAAATGCTTT GGGCACCTTG GATTTAACTG AAAAGTAACC TTGAAACACC GAGGCCCTCA TGCCAGAGGC AAATCGCTGC   -5
  +1
CTCCCGGGTA TTGGCTGCAG CTAAACGGGT CTCTCCAGGC CGACTGAGGT GGGTAGCCAG AAAGAGGACT GAGGTAActg acga
```

C

```
TCTAGAGTTT CCCTCCTTTA GCTGGAGATG ATGAAACACT TAGAGGCAGA AAGATTCCCC GGCTCCCGCT CTGTTGCTCT AGGGCTTTTG -230

GGATCAGCAG GGTGTGGCTC TCTACCCACG CCTTGTAGTC CCCGACTCTT CATCCATCGC AAGCTTCCAG GTGCCAGGAC CGAGGCTCCC -140
                                                                                        SP1
AGTGCCTGAG GTCTCTCTTC TTGCGATCCC GCAGGGGGCG CCCCAATCCG AGGGCGCCGC GCTCGGGGAG GCGCGCGGGG AGCCCGGGGC -50
        TFIID                                            +1
GGGCCCGGGA GGGGGTGTCC CGGCTATAAA AAGTGGCTGC CTCCCAAGGC GCCTGGGCCA gcgggactcc caggctgtgt gcctcaggt
```

FIG. 4. Nucleotide sequence of the 5' flanking regions of the mouse HNF-3α (A), β (B), and γ (C) genes. The transcription initiation sites determined by primer extension analysis (Fig. 3) are shown in boldface, and the 5' most were designated as +1. Potential binding sites for known transcription factors are underlined. The sequences present in the cDNAs are shown in lower case letters.
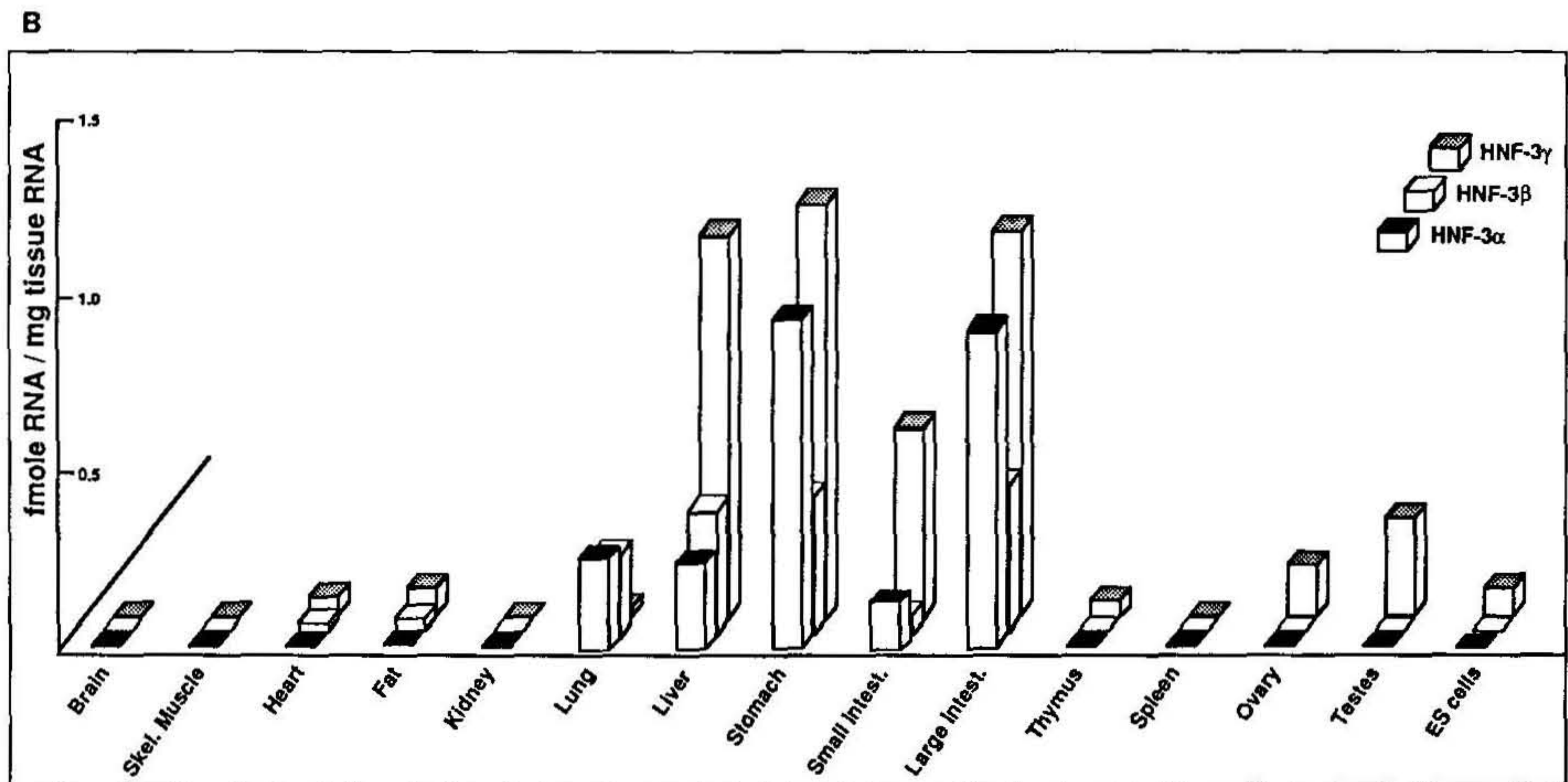
B

FIG. 5. Quantitative analysis of the tissue distribution of the mouse HNF-3α, β, and γ mRNAs. (A) Fifty micrograms of total RNA from the tissues indicated was hybridized to a cocktail of excess ^{32}P-labeled antisense RNA probes specific for HNF-3α, β, and γ, and the level of the mRNAs for all three genes was determined by RNase protection analysis as described under Materials and Methods. The lane marked "Probes" contains 1000 dpm each of the three probes used and served to standardize the values obtained in the quantitation (see below). The arrows indicate the positions of the probes (α, β, and γ) and the protected fragments (HNF-3α, β, and γ). (B) The signals obtained in A were quantified using a phosphoimager and converted to fmol-specific mRNA per milligram total tissue RNA. The signals were corrected for the specific activity of the different probes and reflect the true ratios of the three mRNAs.

at 54°C in 80% formamide overnight. Excess probes were removed by digestion with RNases A and T1 and, the protected probe fragments analyzed on denaturing 6% polyacrylamide gels. The signals obtained were quantified on a Molecular Dynamics phosphoimager and converted to fmol RNA/mg total RNA assuming 100% hybridization of the target RNAs and normalizing the signals for the number of UMP residues incorporated in the hybridizing portion of each probe.

RESULTS AND DISCUSSION

Using a 300-nt probe spanning the forkhead DNA binding domain of the mouse HNF-3α gene (Kaestner *et al.*, 1993), we screened two mouse liver cDNA libraries. A total of 15 phages (10 for HNF-3α, 1 for HNF-3β, and 4

FIG. 6. Temporal expression of the HNF-3α, β, and γ mRNAs during mouse embryogenesis. Twenty micrograms of total RNA from the developmental stages indicated (in days post coitum) were analyzed by RNase protection as described in the legend to Fig. 5. The antisense probes and protected fragments are indicated by arrows.

for HNF-3γ) were purified from 1.3 million phages under high-stringency conditions. The phage inserts ranged from 1 to 2.1 kb and contained the complete coding regions of all three genes. As demonstrated by sequence analysis and transcriptional start site mapping, the cDNA lack less than 50 bp at the 5' end. The cDNAs for HNF-3β and γ are almost full length, as the corresponding mRNAs were shown to be 2.3 and 2.1 kb in size (data not shown). The cDNAs for HNF-3α are lacking 1.6 kb of untranslated sequence at the 3' end, as its mRNA was found to be 3.4 kb in length (data not shown). The sequence of the mouse HNF-3 cDNAs was determined and is shown in Fig. 1 along with the amino acid sequence for the longest open reading frame. The cDNAs encode proteins of 468 (HNF-3α), 459 (HNF-3β), and 353 (HNF-3γ) amino acids with calculated molecular weights of 48.9, 48.5, and 34.6 kDa, respectively. As expected, sequence comparison demonstrated that the mouse and rat sequences (Lai et al., 1990, 1991) are highly conserved, with amino acid similarity ranging from 93% (for HNF-3γ) to 99% (for HNF-3β). A noteworthy exception is the amino terminus of the HNF-3 γ gene, at which the mouse sequence is more similar to the mouse HNF-3α and β genes (16 of the first 55 amino acids are conserved) than to its rat counterpart (6 of 55 conserved amino acids). This is especially interesting in light of the finding that the amino-terminal 52 amino acids of the HNF-3β protein were shown to be important in transcriptional activation through testing of deletion mutants in transfection assays (region IV; Pani et al., 1992a). This domain is rich in serine and tyrosine residues and contains two putative casein kinase I phosphorylation sites. Therefore, it seems likely that the amino terminus of the mouse HNF-3γ gene can function in transcriptional activation as well. This region of the mouse HNF-3γ sequence was also confirmed by sequencing the corresponding genomic region (see below).

Using the mouse cDNA probes for HNF-3α, β, and γ, we screened two mouse genomic phage libraries to ob-

tain genomic fragments containing all three genes. The resulting phages were mapped and the exon-containing regions identified by Southern blotting and subcloning. The restriction patterns of the cloned regions were compared to those obtained in mouse genomic Southern blots to exclude possible rearrangements of the lambda phages (data not shown). The resulting gene structures for HNF-3α, β, and γ are shown in Fig. 2. All three genes span less than 10 kb and contain two (HNF-3α and γ) or three exons (HNF-3β). The intron–exon boundaries were determined by comparing the cDNA and genomic sequences and are summarized in Table 1. Interestingly, the intron within the coding region is at the same relative position in all three genes, that is, at −1 with respect to a conserved YS dipeptide, indicating a common ancestor gene. In addition, the above-mentioned divergence of the amino terminus of the mouse HNF-3γ gene from its rat counterpart cannot be explained by alternate exon usage, as this region covers the first 55 amino acids, while the intron is located at position 24. In addition, no evidence for alternate splice variants of HNF-3γ was found in RNase protection experiments using probes covering the first 452 nucleotides of the HNF-3γ cDNA (data not shown). The intron–exon boundaries of the rat HNF-3β gene, the only other HNF-3 gene analyzed so far (Pani et al., 1992b), are identical to the ones for the mouse gene described here. To determine the start site of transcription for HNF-3α, β, and γ, we performed primer extension experiments with gene-specific primers. The results of these experiments using poly(A⁺) RNA from stomach (where the HNF-3 gene family is expressed most strongly, see below) or tRNA as negative control are shown in Fig. 3 alongside a sequencing ladder using the same primer. In all three cases, multiple, but closely spaced, start sites are apparent. The most 5' of these transcription initiation sites was designated as position +1 (see Fig. 4). The transcriptional start site of the rat HNF-3β gene (Pani et al., 1992b) has been mapped to within 5 nt of the one observed in mouse, again underscoring the close relationship between the two genes.

As an initial step toward the understanding of the regulatory elements governing the expression of the HNF-3 gene family, we subcloned and sequenced the proximal promoters of the three mouse genes and searched them for potential binding sites of known transcription factors (Fig. 4). Despite the relatedness of the three cDNAs and the similarity of the mRNA distribution of the three genes, the promoters are dissimilar. The promoters of HNF-3α and γ have in common only binding sites for the general transcription factor Sp1 (Fig. 4), while only HNF-3γ contains a canonical TATA-box at the appropriate position relative to the start site of transcription. The proximal promoters of the rat and mouse HNF-3β genes are remarkably well conserved, with 116 of the first 120 nucleotides being identical. Especially important is the fact that the two binding sites for a liver-specific protein, termed LF-H3β, and a binding site for HNF-3 itself are conserved in sequence as well as spac-

ing (Fig. 4; and Pani *et al.*, 1992b). Pani *et al.* (1992b) have shown through transfection of rat HNF-3β promoter mutants in HepG2 cells that both binding sites are important for strong expression of chimeric HNF-3β/reporter gene constructs. They proposed a model for cell-specific transcription of the HNF-3 gene in liver cells involving activation by the cell-specific LF-H3β protein and maintenance of this expression through positive autoregulation by HNF-3β. Interestingly, a second HNF-3 binding site of identical sequence is found farther upstream (position −400 to −390) in the mouse HNF-3β promoter, which might also take part in this proposed autoregulation. A second possibility is the regulation of the HNF-3β promoter by HNF-3α and γ, which are also present in liver and have similar DNA binding properties (Lai *et al.*, 1991). Most likely, similar regulatory pathways are operative in the regulation of the mouse and rat HNF-3β genes. This notion is supported by the finding that the proximal mouse HNF-3β promoter confers cell-type-specific expression on a β-galactosidase reporter gene (data not shown). A detailed promoter analysis of the divergent HNF-3α and γ genes through transfection as well as transgenic mouse experiments is in progress to understand which regulatory pathways control these genes, as their proximal promoters lack binding sites for HNF-3 and LF-H3β.

As a first step toward understanding the full potential of the HNF-3 gene family as transcription factors in mice, we studied the expression patterns of its three members through quantitative RNase protection using gene-specific probes (see Materials and Methods for details), allowing for a first direct comparison of the expression levels of the three genes. The results of our survey of adult mouse tissues are shown in Fig. 5. While HNF-3α is expressed strongly in large intestine and stomach, and weaker in liver, lung, and small intestine, HNF-3β mRNA is present in the same tissues at more balanced levels. HNF-3γ is expressed strongly in large and small intestine, stomach, and liver, but is absent from lung. This is consistent with the observation of Monaghan *et al.* (1993) that HNF-3γ expression in midgestation mouse embryos is restricted to derivatives of the endoderm posterior to the liver. The expression of HNF-3α and β in the adult mouse is restricted to derivatives of the embryonic endoderm, while HNF-3γ mRNA was also found in heart, adipose tissue, ovary, and testis as well as embryonic stem cells. The tissue distribution of the HNF-3 mRNAs in mice is qualitatively similar to that observed in the rat (Sladek and Darnell, 1992; Lai *et al.*, 1991). The expression of HNF-3γ in heart, adipose tissue and embryonic stem cells has not been observed before. We extended our analysis of the expression of the HNF-3 gene family to RNAs obtained from mouse embryos from Day 9.5 post coitum onward using the same gene-specific probes (Fig. 6). HNF-3β is expressed strongly on Days 9.5 and 10.5, weakly detected during the next 2 days, and returns to an intermediate level of expression on Day 15.5. HNF-3α follows a similar pattern as HNF-3β, but is expressed at lower levels. The

transient dip in the mRNA levels of HNF-3α and β corresponds to the disappearance of the two mRNAs from the embryonic liver as revealed by *in situ* hybridization (Monaghan *et al.*, 1993). HNF-3γ, while hardly detectable on Day 9.5, is being activated on Day 12.5 and peaks on Days 15.5 and 16.5. This peak in mRNA levels coincides with the increased expression of HNF-3γ in endoderm-derived structures such as the liver, stomach, intestine, and pancreas observed by *in situ* hybridization (Monaghan *et al.*, 1993).

As an initial step toward understanding the role of the HNF-3 gene family of transcription factors, we have cloned and characterized the cDNAs and genes of its three members. The information obtained will allow identification of important regulatory elements in the promoters of HNF-3α, β, and γ through transfection of chimeric promoter/reporter gene constructs into tissue culture cells and by analysis of these constructs in transgenic mice. Furthermore, gene targeting experiments in mouse embryonic stem cells that will allow us to determine the precise role of these genes in mouse embryonic development and tissue-specific transcriptional regulation in the adult are in progress.

REFERENCES

Auge-Gouillou, C., Petropoulos, I., and Zakin, M. M. (1993). Liver-enriched HNF-3 alpha and ubiquitous factors interact with the human transferrin gene enhancer. *FEBS Lett.* **323:** 4–10.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18:** 5294–5299.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81:** 1991–1995.

Clevidence, D. E., Overdier, D. G., Tao, W., Qian, X., Pani, L., Lai, E., and Costa, R. H. (1993). Identification of nine tissue-specific transcription factors of the HNF-3/forkhead DNA binding domain family. *Proc. Natl. Acad. Sci. USA* **90:** 3948–3952.

Costa, R. H., Grayson, D. R., and Darnell, J. E. (1989). Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and α-1 antitrypsin genes. *Mol. Cell. Biol.* **9:** 1415–1425.

De Simone, V., and Cortese, R. (1991). Transcriptional regulation of liver-specific gene expression. *Curr. Opin. Cell Biol.* **3:** 960–965.

Dirksen, M. L., and Jamrich, M. (1992). A novel, activin-inducible, blastopore lip-specific gene of *Xenopus laevis* contains a *forkhead* DNA binding domain. *Genes Dev.* **6:** 599–608.

Frischauf, A. M. (1987). Construction and characterization of a genomic library in λ. *Methods Enzymol.* **152:** 190–199.

Häcker, U., Grossniklaus, U., Gehring, W. J., and Jäckle, H. (1992). Developmentally regulated *Drosophila* gene family encoding the fork head domain. *Proc. Natl. Acad. Sci. USA* **89:** 8754–8758.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III

creates targeted breakpoints for DNA sequencing. *Gene* **28**: 351–355.

Herbst, R. S., Nielsch, U., Sladek, F., Lai, E., Babiss, L. E., and Darnell, J. E. (1991). Differential regulation of hepatocyte-enriched transcription factors explains changes in albumin and transthyretin gene expression among hepatoma cells. *New Biol.* **3**: 289–296.

Hooper, M., Hardy, K., Handyside, A., Hunter, S., and Monk, M. (1987). HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**: 292–295.

Ip, Y. T., Poon, D., Stone, D., Granner, D. K., and Chalkley, R. (1990). Interaction of a liver-specific factor with an enhancer 4.8 kilobases upstream of the phosphoenolpyruvate carboxykinase gene. *Mol. Cell. Biol.* **10**: 3770–3781.

Jürgens, G., and Weigel, D. (1988). Terminal versus segmental development in the Drosophila embryo: The role of the homeotic gene *forkhead. Roux's Arch. Dev. Biol.* **197**: 345–354.

Kaestner, K. H., Lee, K.-H., Schlöndorff, J., Hiemisch, H., Monaghan, A. P., and Schütz, G. (1993). Six members of the mouse forkhead gene family are developmentally regulated. *Proc. Natl. Acad. Sci. USA* **90**: 7628–7631.

Kaestner, K. H., Ntambi, J. M., Kelly, T. J., and Lane, M. D. (1989). Differentiation-induced gene expression in 3T3-L1 preadipocytes: A second differentially expressed gene encoding stearoyl-CoA desaturase. *J. Biol. Chem.* **264**: 14755–14761.

Knöchel, S., Lef, J., Clement, J., Klocke, B., Hille, S., Köster, M., and Knöchel, W. (1992). Activin A induced expression of a *fork head* related gene in posterior chordamesoderm (notochord) of *Xenopus laevis* embryos. *Mech. Dev.* **38**: 157–165.

Kuo, C. J., Conley, P. B., Chen, L., Sladek, F. M., Darnell, J. E., and Crabtree, G. R. (1992). A transcriptional hierarchy involved in mammalian cell-type specification. *Nature* **355**: 458–460.

Lai, E., and Darnell, J. E. (1991). Transcriptional control in hepatocytes: A window on development. *Trends Biochem. Sci.* **16**: 427–430.

Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, H. R., and Darnell, J. E., Jr. (1990). HNF3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes Dev.* **4**: 1427–1436.

Lai, E., Prezioso, V. R., Tao, W., Chen, W. S., and Darnell, J. E., Jr. (1991). Hepatocyte nuclear factor 3α belongs to a family in mammals that is homologous to the *Drosophila* homeotic gene *forkhead. Genes Dev.* **5**: 416–427.

Li, C., Lai, C., Sigman, D. S., and Gaynor, R. B. (1991). Cloning of a cellular factor, interleukin binding factor, that binds to NFAT-like motifs in the human immunodeficiency virus long terminal repeats. *Proc. Natl. Acad. Sci. USA* **88**: 7739–7743.

Li, C., Lusis, A. J., Sparkes, R., Tran, S.-M., and Gaynor, R. (1992). Characterization and chromosomal mapping of the gene encoding the cellular DNA binding protein HTLF. *Genomics* **13**: 658–664.

Li, J., and Vogt, P. K. (1993). The retroviral oncogene *qin* belongs to the transcription factor family that includes the homeotic gene fork head. *Proc. Natl. Acad. Sci. USA* **90**: 4490–4494.

Luckow, B., Bunz, F., Stillman, B., Lichter, P., and Schütz, G. (1994). Cloning, expression and chromosomal localization of the 140 kDa subunit of replication factor C from mouse and man. *Mol. Cell. Biol.,* in press.

McPherson, C. E., Shim, E.-Y., Friedman, D. S., and Zaret, K. S. (1993). An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. *Cell* **75**: 387–398.

Monaghan, A. P., Kaestner, K. H., Grau, E., and Schütz, G. (1993). Post-implantation expression patterns indicate a role for the mouse forkhead/HNF-3 α, β and γ genes in determination of the definite endoderm, chordamesoderm and neuroectoderm. *Development* **119**: 567–578.

Nitsch, D., Boshart, M., and Schütz, G. (1993). Activation of the tyrosine aminotransferase gene is dependent on synergy between liver-specific and hormone-responsive elements. *Proc. Natl. Acad. Sci. USA* **90**: 5479–5483.

Oliver S. G., *et al.* (1992). The complete DNA sequence of yeast chromosome III. *Nature* **357**: 38–46.

Pani, L., Overdier, D. G., Porcella, A., Qian, X., Lai, E., and Costa, R. H. (1992a). Hepatocyte nuclear factor 3β contains two transcriptional activation domains, one of which is novel and conserved with the *Drosophila* fork head protein. *Mol. Cell. Biol.* **12**: 3723–3732.

Pani, L., Qian, X., Clevidence, D., and Costa, R. H. (1992b). The restricted promoter activity of the liver transcription factor hepatocyte nuclear factor 3β involves a cell-specific factor and positive autoregulation. *Mol. Cell. Biol.* **12**: 552–562.

Ruiz i Altaba, A., and Jessell, T. M. (1992). *Pintallavis*, a gene expressed in the organizer and midline cells of frog embryos: Involvement in the development of the neural axis. *Development* **116**: 81–93.

Ruppert, S., Boshart, M., Bosch, F. X., Schmid, W., Fournier, R. E. K., and Schütz, G. (1990). Two genetically defined trans-acting loci coordinately regulate overlapping sets of liver-specific genes. *Cell* **61**: 895–904.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463–5467.

Sasaki, H., and Hogan, B. L. M. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development* **118**: 47–59.

Sawaya, P. L., Stripp, B. R., Whitsett, J. A., and Luse, D. S. (1993). The lung-specific CC10 gene is regulated by transcription factors from the AP-1, octamer, and hepatocyte nuclear factor 3 families. *Mol. Cell. Biol.* **13**: 3860–3871.

Sladek, F. M., and Darnell, J. E. (1992). Mechanisms of liver-specific gene expression. *Curr. Opin. Genet. Dev.* **2**: 256–259.

Tao, W., and Lai, E. (1992). Telencephalon-restricted expression of BF-1, a new member of the HNF-3/*fork head* gene family in the developing rat brain. *Neuron* **8**: 957–966.

Weigel, D., and Jäckle, H. (1990). The *fork head* domain: A novel DNA binding motif of eucariotic transcription factors. *Cell* **63**: 455–456.

Weigel, D., Jürgens, G., Küttner, F., Seifert, E., and Jäckle, H. (1989). The homeotic gene *fork head* encodes a nuclear protein and is expressed in the terminal regions of the drosophila embryo. *Cell* **57**: 645–658.