The Homeodomain Resource: sequences, structures, DNA binding sites and genomic information

Sharmila Banerjee-Basu, Daniel W. Sink and Andreas D. Baxevanis*

Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

Received October 3, 2000; Accepted October 25, 2000

ABSTRACT

The Homeodomain Resource is an annotated collection of non-redundant protein sequences, three-dimensional structures and genomic information for the homeodomain protein family. Release 3.0 contains 795 fulllength homeodomain-containing sequences, 32 experimentally-derived structures and 143 homeobox loci implicated in human genetic disorders. Entries are fully hyperlinked to facilitate easy retrieval of the original records from source databases. A simple search engine with a graphical user interface is provided to query the component databases and assemble customized data sets. A new feature for this release is the addition of DNA recognition sites for all human homeodomain proteins described in the literature. The Homeodomain Resource is freely available through the World Wide Web at http:// genome.nhgri.nih.gov/homeodomain.

INTRODUCTION

The homeodomain is a common DNA-binding structural motif found in many eukaryotic regulatory proteins (1,2). Homeodomain proteins are involved in the transcriptional control of many developmentally important genes, and 143 human loci have been linked to various genetic and genomic disorders. X-ray crystallographic and NMR spectroscopic studies (3–11) on several members of this family have revealed that the homeodomain motif is comprised of three α -helices that are folded into a compact globular structure. Helices I and II lie parallel to each other and across from the third helix. This third helix is also referred to as the 'recognition helix', as it confers the DNA-binding specificity of individual homeodomain proteins. The homeodomain has been evolutionarily conserved at the structural level (12); this is most evident upon examination of divergent members of the homeodomain family.

The Homeodomain Resource represents a comprehensive collection of information about the homeodomain family. The database contains all available full-length and homeodomain-only sequence data and structures as of October, 2000. The genetic data contained in this database includes information on human diseases in which homeodomain-containing proteins are implicated, cytogenetic map locations and specific mutation data underlying the disease condition. Since its last release

(13), 27 new loci for genetic disorders have been identified. The sequence information within the database is automatically updated on a monthly basis, with each entry in the database rigorously selected to assure non-redundancy.

DATABASE DESCRIPTION

The current version of the database contains 795 full-length homeodomain protein sequences isolated from 83 different species (Table 1). The complete full-length sequence data as well as the homeodomain portion of the sequence is available in FASTA format. The database can be searched on the basis of SWISS-PROT ID, GenBank accession number, gene names (both common and alternative), protein description, sequence and organism name. The search engine also supports Boolean queries, allowing users to search on individual fields or on all fields at once. Search results are returned in a tabular format, with hyperlinks to the original records in GenBank and SWISS-PROT, respectively. Individual sequences can also be retrieved in FASTA format from a pop-up window.

Table 1. Homeodomain Resource statistics

Total sequences available	3796	
Non-redundant full-length sequences	795	
Genes/gene symbols	375	
Distinct organisms	83	
Three-dimensional structures	32	
Homeobox loci implicated in human genetic disorders	143	

The genetic information available for the homeodomain protein family has increased ~23% since the last release, with the inclusion of 27 new loci. The genetic data are compiled from both the literature and from the Online Mendelian Inheritance in Man (OMIM) database at NCBI (http://www.ncbi.nlm.nih.gov/Omim/). As before, a search engine is available for querying this genomic information. Search results are presented in a tabular format with hyperlinks to the original records and can be sorted by disease name, map location, gene symbol, protein name or OMIM identifier. This value-added format allows users easy access to related information.

*To whom correspondence should be addressed. Tel: +1 301 496 8570; Fax: +1 301 402 6858; Email: andy@nhgri.nih.gov

ARIK G T C A A T T A G OCT 1 A A A T G C A A T A C C MAX2 Y Y A A T G A T T A Y G OCT 2 A T G C A A T C CDX1 C T T T A T G OCT 2 A T G C A A T C CDX2 C T T T A T G PA2 C C T A A T C C G G G C T A A T C CDX4 C T T T A T C PA2 T C A T G C G C T A T T G A C C T CDX4 T T A T C C C PA3 T C A T G C G C T A T T G C C C CX T A A T C C G PA3 T C A T G C C T A T C C C T CX T A A T C C G PA3 T C A T T C C C T A T C C C T CX T A A T T C A T C C G PA3 A H H T T C C C C T A T C C T DX3 C C C T A T T C A G PA3 A H H T T C C C C T A T C C T DX4 C C A T A T T A C G T PA3 T C A T T C A T C C T DX4 C C T A A T T C A G PA3 T C A T T C C T C A T C C T DX4 C A T A T T A C G T F A T T C A G PA3 T C A T T C A T C C T T A A T T C A T C A T T C A T C A T T C	Common Name	DNA Binding Sequence	Common Name	DNA Binding Sequence
BARK2 Ý Ý TA AT G A T T T T Y OCT 2 A T G G A A T COX1 C T T T A T G OTT 2 C C T A A T C C T G G G G T T A T G COX2 C T T A A T G PAZ T G A T C C T G G G G T T A T G COX4 P T A A T G PAZ T G A T C C T G G G G T T A T G CRX T A A T G A T G PAZ T G A T C C T G G G G T T G A C CRX T A A T G A G G PAX A H H M T G A C C C T G G G T G A T G C CXX T N A A G T G PAX A H H M T G A C C C T G G G T G A T G C DX2 C A C G A T T A T G A G G PAX A H H M T G A C C T T G A T G C DX3 C A C G A A T A T G G A G PAX A A H H M T G A C C T T G A T G C DX4 C A A T A T G G A G PAX A A H H M T G A C C T T G A T G C DX5 C A A T A T T A G PAX A T A D T T A G A T G A A T G A T T A A T G A G A	ARIX	GT CAAT TAG	OCT 1	AATGCAAATACC
CDXLC T T T A T GOCT3A T G C A A T CCDX2C T T T A T GPAX2C C T A A T C C G G G G T T A T CCDX4C T T A A T CPAX2T C A T C C G G G G T T A T CCR1T A A T C A T C C CPAX3T A A T L B A T ACRXT A A T C A C C TPAX3T A A T L B A T ACRXC C C T A A T C A T C C T C C TTCDX3C A C C T A T T A A T C APAX3G C A C C C T A C C TCDX3C A C T A T T A A T C APAX3T C A C C T A T C C TDX4C A C T A T T A A T C APAX3T C A C C T A T C C T CDX5C A C T A T T A T T C APAX1T A T C A T T C A T C C TDX4C C A T T T A T T C APAX1T C A T C C T T T T C C T C T CDX5C A T A T T A C APAX1T C A T C C T T T C C T C T C T C T C T	BARX2	YY TAAT GRTTTTY	OCT 2	ATGCAAAT
CDX2CCTTTTATCPAX2CCTTTTATCCDX4TTTATTPAX2TCATCCTCCCTCCTTCCDX4TTTATTPAX2TCATCCTCCCTCCTCCTCCCTCCCTCCCTCCCCCCCCCC	CDXI	CT TTAT G	OCT3	ATGCAAAT
COTA CONTACT OF CONTAC	CDX2	AAA	OTX2	
CLUX TEATENT PA2 TEATENT A AT NET A TEATENT CA CRX TAATEN ATTA TEATENT A AT NET A TEATENT AT A AT NET AT TAATEN A TEATENT A CRX TAATECA PAXA A HIN TECA C C C CX TAATECA PAXA A HIN TECA C C C CX TAATECA PAXA A HIN TECA C C C T CX TAATECA PAXA A HIN TECA C C C T DLX C A C TAATECA PAXA A HIN TECA C C C T DLX C A C TAATECA PAXA TAATE DLX C A C TAATECA PAXA TAATE DLX C A C TAATECA PAXA TAATE DLX C A C TAATECA PAXA TAATECA C T DLX C A C TAATECA PAXA TAATECA C PAXA TAATET DLX C A TAATETA PA PAXA TAATECA TAATET DLX C A TAATETA PA PAXA TAATETA PAXA TAATETA DLX A C A AT AATETA PAXA PAXA TAATETA PAXA PAXA PAXA PAXA PAXA PAXA PAXA PA		A		CCTAATCCTGGGTTATC C G
CHITAATENIMATTAPAGITAATEACKXTAATEAPAGIAN HINTCACCCTCSXTHAAGTOPAGIG T C A C C C T C A T CTDAGIC A C T C A T T C A G OPAGIG T C A C C C O C A T C C A T CDAGIC A C T A A T C A G OPAGIA A HINT T C C C G O T C A T C C A CDAGIC A C T A A T C A G OPAGIT T C A T T A T C A T C A C C T A T C A C C T A T C A C C T A T C A C C T A T C A C C T A T C A C C T A T C A C C A T C A C C A T C A C C A T C A C C A T C A C A	CDX4	TTATT	PAX2	TCATGCATGAC
CKX TAATCO CXX TAATCO CXX TAATCO CXX TAATCO CXX TAATCO CAC TAATTO A DU2 CAC TAATTO A C C C T C AT C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C AT C C C C C C C T C C T C C C C C C C C T C C T C C C C C C C C T C C T C C C C C C C C T C C T C C C C C C C C T C C T C C C C C C C C C T C C T C C C C C C C C C T C C T C C C C C C C C C T C C T C C C C C C C C C T C C T C C C C C C C C C T C C T C C C C C C C C C T C C T C C C C C C C C C C T C C C T C C C C C C C C C C T C	CRT1	TAATNNNATTA	PAX3	TAATNNATTA
CX THAAGTG PAX PAX C T CACT CAC T DX2 CACTAATTGAC PAX C A C T CATGATC C DX3 CANTAATTAATCGAC PAX C A C TAATTCAC DX3 CANTAATTAATCGAC PAX C A T CATTAATT DX3 A CATTAATTGA PAX C A T CATTAATCAC DX3 CANTAATTAAC PEXIHOXA9 T CATTAATT DX4 A TAATTAC PEXIHOXA9 T CATTACT DX4 C CACTAATTGAC PEXIHOXA9 T CATTACT DX4 C CACTAATTGAC PEXIHOXA9 T CATTACT DX4 C CACTAATTGAC PEXIHOXA9 T CATTAC EAX C C C C T A T T A T T N N N T A A T N N N N T A T T N EAX C C C C T A T T C C C PEXIHOXA9 T C C T T C C EAX C C C C T A T T C C C PEXIHOXA9 T C C T T T C EAX C C C C T A T T C C PEXIHOXA9 T C C T T T C C EAX C C C T T A T T C C PEXIHOXA9 T C C T T T C C EAX C C C T T A T T C C PEXIHOXA9 T C C T T T T C C C T T HX1 C T A T T A T T C C C PEXIHOXA9 A T C A T T A T C C T C T A HX1 C T A T T A T T C C C PEXIHOXA9 A T C A T T A T C C T T T HX1 C T A T T A T T C C C PEXIHOXA9 A T C A T T A T C C T T HX1 C T A T T A T T C C C PEXIHOXA9 A T C A T T A T C C T T HX1 C T A T T A T T C C C PEXIHOXA9 A T C A T T A T C C T T HX1 C T A T T A T T C C C T PEXIHOXA9 A T C A T T A T C C T T HX1 C T A T T A T T C C C C PEXIHOXA9 A T C A T T A T C C T T HX1 C T T T T A T A T T C C C T T HX1 C T T T T T C A C C C PEXIHOXA9 T C C T T C A T T C C T T HX1 C C T T T T T C A C C C PEXIHOXA9 T C C T T C A T T C C T T HX1 C C T T T T T C A C C PEXIHOXA9 T C C T T T A T T A C C T T HX1 C C T T A T T C A C C PEXIHOXA9 T C C T T T A T T A C C T T HX1 C C T T A T T C A C C PEXIHOXA9 T C C T T T A T T A T C C T T HX1 C C T T A A T T C C PEXIHOXA9 T C C T T T A T T A T C C T T HX1 C C T T A A T T C C C PEXIHOXA9 T C A T T A T C C T T T HX1 C T T A C C A C P PAXIH T A T C C T T T A T C C A C T T HX1 C T T A C C A C C PAXIT T A T C C A C T T T A T C C A C T T HX1 C T T A C C A C C PAXIT T A T C C A C T T A T C C A A T T T C C A C T T T A C C A C T T T T	CRX	С таатса	PAX4	
CAX T MAX 0 T G PAS G C C C T T G A T G C DX2 C A C T A A T T G C G G PAXG A A H H T C A C G C T T G A T G C A C DX3 C A C T A A T T G C A G PAXG T A A T T A DX4 C A A T T A A T T G A PAXG T A A T T DX5 C A A T T A A T T G A PAXG T G A T T G A T G C A T G C A C DX4 T C A A T T A A T T G A PAXG T G A T T G A T G A T DX5 A T A A T T A T G A PAXG T G A T T A T G A T DX4 A T T A A T T A G PAXG T G A T T A T G A T DX5 C A C T A A T T A G A PAXG H A T T A G A T T A T A G A T T A T A G A DX4 A T G A T T A T A T A T G A PAXG H A T G A G A T T A T T A T G A T G A T T A T A	CEV	I R R I C R	DAVE	A C T
DX2 C A C T A A T T A T C A C PAX0 A A H N T T C A C C T A T A C C T A T DX3 C A H T A T T C A A T T A T T C A T T A T T C A T PAX1 T C A T T A T T C A T T A T T C A T DX4 T C A A T T A A T T C A T T A T T C A T T A T T C A T T A T T C A T T A T T C A T T A T T C A T T A T T C A T T A T T C A T T A T T C A T T A T T C A T T A T C A T T A T T A T T C A T T A T T A C T T A T C A T T A T T A C A T T A T T A T C A T T A T A	CSA	TNAAGTG	РАХЭ	G TCAC GCTTGATGC
DX3 CANTATATORA PAT ATTATORA PAT TATATATATATATATATATATATATATATATATA	DLX2	Састаатисас	PAX6	Α С ΤΑ Τ ΑΝΝΤ ΤΟΑΟΘΟ ΤΤΟΛΟΟΟ
DX4 T C A A T T A A T T G A PX1 T C A A T T A A T T G G DX5 A T A A T T A G PX1.HOXA9 T G A T T T A T G A DX6 A T A A T T A G PX1.HOXA9 T G A T T A T G A EAX2 C A C T A A T T A T T G A PX1.HOXA1 A T G A T T G A T G A C G A GX2 A O T G A G T A A T T G A PX1.HOXA1 A T G A T T A T G G A GX2 A O T G A G T A T T A T T G G PX1.HOXA1 A T G A T T A T T G O A IX1 T A T T A T T G G PX1.HOXA5 A T G A T T T A T G C T C T A IX2 C G C T T A T T A T T G A PX1.HOXA5 A T G A T T T A T G C T C T A IX1 T C T A T T A T T G A PX1.HOXA5 A T G A T T T A T G C T T T A T G C T T A IX1 T C T A T T A T T G A PX1.HOXA5 A T G A T T T A T G C T T T A T T G C C T T A IX1 T C T A T T A T T G A PX1.HOXA5 A T G A T T A T T G C T T T T G C C T T A IX11 C G C T T G G C A A A G G T C A C C T PX1.HOXA5 G T G A T T A T C C T T T A T G C T T T T T G C T T T T T T G C T T T T	DLX3	CAN TTAT CTCAG	PAX7	TAATT
DXS ATAATTAG PEXLHOXAN T G ATTAT G G DXG ATAATTAG PEXLHOXAN T G ATTATC DXG C AC TAATTAG PEXLHOXAN T G ATTAC DXG C AC TAATTAG PEXLHOXAN AT G ATTAC DXG C AC TAATTAG PEXLHOXAN AT G ATTAC DXG C AC TAATTAG PEXLHOXAN AT G ATTATCG DXG A G T A ATTAG PEXLHOXAN AT G ATTATAG C ATTAC DXG C AC TAATTAG PEXLHOXAN AT G ATTATCG C ATTAC HXI T C ATTAATTG C A PEXLHOXAN AT G ATTATCG C ATTACC HXI T C T G TAATTAC A PEXLHOXAN AT G ATTATTCG C ATTACC HNEI T C T G TAATTAC A PEXLHOXAN AT G ATTATTCG C ATTACC HNN C G T TAATTAC T T C A PEXLHOXAN T T G ATTATC C T T HNN C G T TAATTAC T T C T PEXLHOXAN T T G ATTATC C T T HNN C G T T AT T T T T T T T T T T T T T T T	DLX4	Τ C A A T T A A T T G A	PBX1	T T G A T T G A T
DLX6 A T A N T T T A C PEXLHOXA9 T G A T T T A T EXY2 C A C T T T A T T T G A G PEXLHOXA1 A T G A T T T A T C GBX2 A G T G A G T A T T A T T T G A C PEXLHOXA1 A T G A T T A T C A GBX2 A G T G A G T A T T A T T T G G A PEXLHOXA1 A T G A T T A T T A T C A HX T A A G T A T T A T T A T G A PEXLHOXA9 A T G A T T T A T G C C T T A HX1 T A T T A T T A T C A PEXLHOXA9 A T G A T T T A T G C C T T A HX1 T A T T A T T A T C A PEXLHOXA9 A T G A T T T A T G C C T T A HX1 T A T T A T T T G A C A PEXLHOXA9 A T G A T T T A T G C T T A HX1 T A T T A T T T G A C A PEXLHOXA9 A T G A T T T A T G C T T A HX1 T C T G G T A A T T A C A PEXLHOXA9 A T G A T T T A T T C C T T HX1 C G T T G G C A A G G T C A C C T PEXLHOXC6 G T G A T T T A T T G C T T HX1 C G T T A A T T G G PEXLHOXC6 G T G A T T T A T G C T T HX1 C G T T A A T T G G PEXLHOXC6 G T G A T T T A T G C T T HX1 C G T T A A T G G PEXLHOXC6 G T G A T T A T G C T T HX11 C G T T A A T G G PEXLHOXC6 G T G A T T A T G C T T HX11 C G T T A T T G G PEXLHOX	DLX5	A T A A T T A G	PBX1-HOXA10	A T G A T T T A T G A
EM2 C A C T A A T T G A G PEXI-HOX81 A T G A T T G A T C G A EVAI T N A T T A N N N N N T A A T G PEXI-HOX86 A T G A T T A T T G A GEX2 A G T G A G T A A T T G G PEXI-HOX86 A T G A T T A T A G A HEX T A A T T A A T T A A T T G A PEXI-HOX86 A T G A T T T A T G A C T C T A HK31 T A A T T A A T T G A PEXI-HOX86 A T G A T T A T A C T G A HM21 T C T G T A A T T A C A PEXI-HOX86 A T G A T T A T A C T T G C T T A HN81 T C T G T A A T T A C A PEXI-HOX86 A T G A T T A T A T T T G C T T A HN81 T C T G T A A T T A C A PEXI-HOX86 A T G A T T A T A T T T G C T T A HN81 T C T G T A A T T G C C C P PEXI-HOX86 A T G A T T A T A C T T T G C T T A HN84 G G C T T G G C A A A G G T C A T T T G C C T T A T T G A T T A T T T C C T T T T G G C T T A T T G C T T T T G G C T T A T T G C T T T T G G C T T A T T G C T T T T G G C T T A T T G C T T T T G G C T T A T T G C T T T T G G C T T A T T G G C T T A T T G G C T T A A T T G C C T T A T T G C C T T A T T G C C T T A T T G G C T T A A T T G C C P PEXI-HOX80 G T G A T T A A T C C T T T A T G C C T T A A T T G C C P PEXI-HOX80 HN84 C G C T T A A T T G C C P PEXI-HOX60 G T G A T T A A T C C T T A T C C P PEXI-HOX60 G T G A T T A T C C T T A T G C C P PEXI-HOX60 HOX11L1 C G T T A A T G C C P PEXI-HOX60 G T G A T T A T C C	DLX6	АТААТТАС	PBX1-HOXA9	тсаттас
EVX1 TNATTANNNNTAATNO PXX1-HOX86 ATGATTGATGATGA GBX2 AGTGAGGAGTATTGG PXX1-HOX86 ATGATTATGGTCTA HEX CAAG PXX1-HOX86 ATGATTATGGTCTA HKX1 TCTATTATTGACA PXX1-HOX86 ATGATTATGGATCTATGACA HMI TCTGTATTATTGACA PXX1-HOX66 ATGATTATGGACA HNN1 AGTTATTTGGACA PXX1-HOX66 ATGATTATTGGACA HNN1 AGTTATTTTGGACA PXX1-HOX66 GTGATTATTGCACA HNN4 CGCTTGGCAAAGGTCACCT PXX1-HOX66 GTGATTATTGCTT HN76 GATATTGATTTTT PX2 TTGATTATTGCTT HN76 GATATTGACA PXX1-HOX66 GTGATTTATTGCTT HOX111 CGTTAATTGG PXX1-HOX66 GTGATTTATTGCTT HOX112 CGTTAATTGG PXX1-HOX66 GTGATTTATTGCTT HOX111 CGTTAATTGG PXX1-HOX66 GTGATTTATTATTGCTT HOX112 CGTTAATTGG PXX1-HOX66 GTGATTTATTATTGCTT HOX112 CGTTAATTGG PXX1-HOX66 GTGATTTATTATTGCTT HOX112 CGTTAATTGGA PXX2-HOX66 GTGATTTATCCTT HOX112 CGTTAATTGA	EMX2	CAC TAAT TGAG	PBX1-HOXB1	ATGATTGATCG
GEX2 A G T G A G T A A T T G G FPX1-H0X85 A T G A T T T A T T A HEX T C A G T A A T T G G PDX1-H0X85 A T G A T T T A T G G C T C T A HX1 T C T G T A T T A A T T G A PDX1-H0X85 A T G A T T T A C G A C C T A HX1 T C T G T A A T T A C A PDX1-H0X85 A T G A T T T A C G A C C T T HME1 T C T G T A A T T A C A PDX1-H0X86 A T G A T T A C G A C HNF1 A G T T T A T T T G A C A PDX1-H0X86 A T G A T T A C G A C HNF1 A G T T A T A T T C A C A PDX1-H0X66 A T G A T T A C A C T T T HNF4 C G C T T A T T T C A C A PDX1-H0X66 G T G A T T G A T G A T C G T T A T T T G A T T G A T T G A T T G A T G A T T A C T T HNF4 C G C T T A T T T C A C C C T PDX3-H0XC6 G T A T T A T T G C T T H0X111 C G T T A T T G A PDX3-H0XC6 G T G A T T A T T A C T T H0X111 C G T T A T T G A PDX3-H0XC6 G T A T T A T T A C T T H0X111 C G T T A T T G A PDX3-H0XC6 G T A T T A T T A C T T H0X111 C G T T A T T G A PDX1-H0XC6 T A A T G A T T A T T A C T T H0X12 T T A A C G A C POU1F1 <td< td=""><td>EVX1</td><td>T N A T T A N N N N T A A T N G</td><td>PBX1-HOXB4</td><td>A T G A T T G A T G A</td></td<>	EVX1	T N A T T A N N N N T A A T N G	PBX1-HOXB4	A T G A T T G A T G A
HEX C A A C T A A C T A PSXI-HOXE7 A T G A T T A A T G C T C T A HK31 T T A A C T A PSXI-HOXE8 A T G A T T A A T G C T C T A HK31 T C T A T T A A T T G A PSXI-HOXE8 A T G A T T A T G A C T C T A HK51 T C T G T A A T T G A C A PSXI-HOXE9 A T G A T T A T A T G A C T T T HNF1 A G T T T A T A T T T G A C A PSXI-HOXE6 G T G A T T A T A T T G C T T T HNF4 C G C T T G G C A A A G T C A C C T PSXI-HOXE6 G T G A T T A T A T T G C T T T HNF4 C G C T T G G T A A T T G C A C C T PSXI-HOXE6 G T G A T T A T A C T T HNF6 C G C T T A A T T G C A C C T PSXI-HOXE6 G T G A T T A T G C A C G C T HNF6 C G C T T A A T T G C A PSXI-HOXE6 G T G A T T A T G C A C G C T HOX111 C G T T A A T T G A PTX1 T T A A T C C HOX112 C G T T A A T T G A PTX1 T A A T C C HOX114 C G T T A T T A T G A PTX1 T A A T C C HOX112 T T T A T T A T G A C PX1 T A A T C C HOX114 T T T A C A POUFI T A A T C C HOX15 T T T A C G A G<	GBX2	A G T G A G T A A T T G G	PBX1-HOXB6	A TGATTTAT T A
HK31TAATAPEXLHOX88ATGATTATGACTCAHLX1TCTATTATATTGACAPEXLHOX89ATGATTACGACHMEITCTGTATATATTGACAPEXLHOXC6AHNFIAGTTTATATTGACAPEXLHOXC6AHNF4CGCTTGGCAAAGGTCACCTPEXLHOXC6GCGCTTGGCAAAGGTCACCTPEXLHOXC6GGHNF4CGCTTGGCAAAGGTCACCTPEXLHOXC6GGGTTATATTGACAPEXLHOXC6GGHNF4CGCTTAGCCAAAGGTCACCTPEXLHOXC6GGGTTATTGGCPEX2TTGATTGATHOX112CGTTAATTGGPEX2HOXC6GGGTTATGACPEX3HOXC6GGHOXA10TATTGAPITX2TAATCCHOXA9TTTACGACPIX1TAATCCHOXB8TTTAGGACPIX1TAATCCHOXB9TTTACGACPIX1TAATCHOXB9TTTACGACPIX1TAATCHOXD12TTTACGACPIX1TAATGAAATHOXD12TTTACGACPOUSFIATGCAAATHOXD13TTAATGACGPOUSFIAISL1GCTTAAATACGGPOUSFIATGCAAATHI2TTAATGACGGPOUSFIATGCAAATHI2TGACAGTTPOUSFIATGCAAATHI2TGACAGTTPOUSFIATGCAAATHI2TGACAGTTPOUSFIATGCAAATHX11GCTCATTAACGGCTCCATTAACCHX2TCAATTAATGAPOUFIATGCAAATHX11GCACAGTTGCTCCATTAACCHX12TGACAGTTGCTCAT	HEX	T T CAAG	PBX1-HOXB7	ΣΤΓΓΑΤΤΤΑΤ ΩΟΤΟΤΔ
HLXI T C T A T T A A T T G A PEXL-HOXEG A T G A T T T A C A A T G A T T T A C A HMEI T C T G T A T T A C A PEXL-HOXEG A T G A T T T A T A T T G G C A HNF4 A G T T T A T A T T G A C A PEXL-HOXEG A T G A T T A T A T T G C T T T HNF4 C G C T T G G C A A A G G T C A C C T PEXL-HEISI T G A T T G A T T G A C G C T T G A T T T A T G A C G A C A T T T T T G A A T T T T G A A T T T T	HK31	ΤΑΑΓΤΑ	PBX1-HOXB8	ATGATTTATGACTCTA
HMEI T C T G T A A T T A C A PEXI-HOXC6 A T G A T T A T T T T T T G A C A HNFI A G T T T A T A T T T G A C A PEXI-HOXC6 G T G A T T A T T T A C T T T HNF4 C G C T T G G C A A A G G T C A C C T T PEXI-HOXC6 G T G A T T A C A C C T T T HNF6 G A T A T T G A T T T T T PEX2 T G A T T G A T G C C T T T HOX111 C G T T A A T T G G PEX2-HOXC6 G T G A T T T A T G C C T T T HOX112 C G T T A A T T G G PEX2-HOXC6 G T G A T T A T T A C C T T T HOX10 T T A T G G PEX2-HOXC6 G T G A T T A T T A C C T T T HOX111 C G T T A A T T G G PEX2-HOXC6 G T G A T T A T A C C T T T HOX10 T T A T G G PEX2-HOXC6 G T G A T T A T T A C C T T T HOX1112 C G T T A A T T G G PEX2-HOXC6 G T G A T T A T T A C C T T T HOXA0 T T A T G G A PIT I T A A T C C HOXB3 T T T A T G G A C PIX1 T A A T C C HOXB9 T T T A C G A C PIX1 T A A T A C G A C HOXD10 T T A A C G A C POUFI T A A A T A C G A C HOXD13 T T A A T G G G G POU3F1 A T G C A A A T ISL2 T T A A T G G G G POU4F1 A C C C A A A T T T	HLX1	Τ C T A T T A A T T G A	PBX1-HOXB9	A T G A T T T A C G A C
INFLAGTTTATATATGACAPSXI-HOXC6GTGATTTATGUTTTHNF1AGTTTATATATTGACAPSXI-HOXC6GTGATTTATGUTTTATHNF4CGCTGGGATAAGGCACCTPSXI-MEISITGATTGACAGCTHNF6GATATTGATTTTTPSX2TTGATTGACAGCTHNF1CGTTAATTGGPEX2-MOXC6GTGATTTATTGCGTTTHOX111CGTTAATTGGPEX3-HOXC6GTGATTTATTGCTTTHOX1112CGTTAATTGGPEX3-HOXC6GTGATTTATGCTTTHOX1112CGTTAATTGAPTX1TAATCCHOXA0TTTATGCACHOXA9TTTAATCCHOXA9TTAGCACCPOUIFIHOXB9T TAAGACPOUIFITAAATAHOXB9T TAACGACPOUIFITAAATAHOXD12TTAACGACPOU3FITGGAGATATACGAGHOXD13T TTACGAGPOU3FITGGAGATATACGAGHOXD13T TAACGACPOU4FIATIFIC T C TAATTGGGPOU4FIG C T C ATTAACISL-1G C T TAATGGGPOU4FIG C T C ATTAACISL-2T TAAGTAPOU4FIG C T C ATTAACIH2T TAATGGAPOU4FIA TMSX1T C AATTAATTGGAPOU4FIA TMSX2T C AATTAATTAGAPOU5FIA TG C A AATMSX2T C AATTAATTAGATGFT AATGAACMSX2T C AATTAATTACATGFT G A C AAMSX2T C AATTAATTACTATGFT G A C AAMSX2T C AATTAATTACATGFT G A C AAMSX2T	HMEI		PBX1-HOXC6	G A
HNP1 À G T T T A T T A T T T G A C A PBX HOX6 G T G A T T A T T A T T A C T T T HNP4 C G C T T G G C A À A G G T C A C C T PBX HOX6 T G A T T G A C G C T HNP6 G A T T T T T T T T T T T T T T T T T T		ICIGIAAIIACA		A TGATTTATTGCTTT A
HNF4 C G C T T G G C A A A G G T C A C C T PEXI-MEISI T G A T T G A G T G HNF6 G A T A T T G G A T T T T T PEX2 T G A T T G A T G HNX111 C G T T A A T T G G PEX3-HOXC6 G T G A T T G A T G HOX111 C G T T A A T T G G PEX3-HOXC6 G T G A T T A T T G C T T G HOX1112 C G T T A A T T G G PEX3-HOXC6 G T G A T T A T T A C C T T HOXA10 T T T T T G A PITX1 T A A T C C HOXA9 T T T T T G A C PMX1 T A A T C C HOXB7 T T T A T G A C PMX1 T A A T A HOXB8 T T T A T G A C PMX1 T A A T A HOXB7 T T T A T G G A C PMX1 T A A T A HOXB7 T T T A T G G A C PMX1 T A A T A HOXB7 T T T A T G G A C POUIFI T A A T A HOXB1 T T T A C G A C POU3F1 T T G C A A A T HOXD12 T T A A T A T C T G POU4F1 G C T C A T T A A C G IFI C T C T A A T A T C T G POU4F2 G C T C A T T A A C G ISL-1 G C T T A A T A T C T G POU4F2 G C T C A T T A A C G ISL2 T T A A T T G A POU4F2 G C T C A T T A A C G	HNFI	A G T T T A T A T T T G A C A	PBX1-HOXC6	G TGATTTAT TACTTT
NF6 GATAT GATATTGATTTT PEX2 T GAT GAT T GAT GAT T GAT GAT T GAT GAT T GAT	HNF4		PBX1-MEIS1	G
HOX II C G T T A A T T G G PEX2 HOX CG FE G A T T A T T G G A G T T T A T T G A C T T T T T G A T T T A T G A C PIXI T T A T T A C C PIXI T A A T C C PIXI T A A A T T A C A A A T T T A T G A C PIXI T T A T G A C A G T T T A T G A C A C PIXI T T A T G A C A G T T T A T G G A C PIXI T A T A C A A A T T T A T G A C A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C PIXI T T A T G C A A A T T T T A C G A C T T T A A C G A C PIXI T T A T G C A A A T T T T A C G A C T T T A A C G A C T T T A A C A A T T T A C G C T C A T T A A C A A A T T T A C G C T C A T T A A C A A A T T T A C G C T C A T T A A T A T C C A A T T A C T T A A T T A C T T A A T A T	HNF6	GATATTGATTTTT	PBX2	TGATTGACAGCT TTGATTGAT
HOXILLI C G T T T A A T T G G PEX3 T T G A T T G A G HOXILL2 C G T T T A A T T G G PEX3-HOXC6 G T G A T T T A T T A C T T T HOXA10 T T T T G A PITXI T A A T C C HOXA9 T T T T A T G A C PMX1 T A A T A HOXB7 T T T A T G A C PMX1 T A A T A HOXB8 T T T A T G A C POUIFI T A A T A HOXB9 T T A A C G A C POUIFI T A A C G A A A T HOXD10 T T A T A C G A C POUSFI A T G G A A A T HOXD13 T T T A C G A G POUSFI A T G C A A A T HOXD13 T T A A A G G G POU4FI T G G C A G A A A ISL-1 G C T T A A T A T A T C T G POU4FI A T G C A A A T ISL-2 T T A A G G T A POU4FI G C T C A T T A A C ISL-2 T T A A G G T A POU4FI G C T C A T T A A C ISL-2 T T A A G G T A POU4FI G C T C A T T A A C ISL-2 T T A A T G T A A G T A T A A C T A A C ISL-2 T T A A T A A C T A A C T A A C <td>HOX 11</td> <td>CGTTAATTGG</td> <td>PBX2-HOXC6</td> <td>GTGATTTATTGCTTT</td>	HOX 11	CGT TAAT TGG	PBX2-HOXC6	GTGATTTATTGCTTT
HOX11L2CGTTAAATTGGPBX3-HOXC6GTGATTTATTACTTTHOXA10TATTGAPITX1TAATCCTTTHOXA0TATTGAPITX1TAATCCTTHOXA9TTTCPITX2TAATCCTTHOXB6TTTATGACPMX1TAATAHOXB7TTTATGACPOUIFITAATAHOXB8TTTATGACPOUIFITAAATHOXD10TTTATGACPOUSFITTAATGAAATHOXD12TTAACGACPOUSFITGGAATTATACGAAGHOXD13TTTATGGAGPOUFIÅISL-1GCTTAAATAGGGPOUSFITGCAAAATISL-2TTAATGAAPOUSFIATTAATGAAPOUFIÅTMX1TGAAGTAAPOUFIAMX2TCAATTAATGGPOUSFITGCAAATMSX1CTAATTGAPOUFIAMSX2TCAATTAATGGAPOUSFITGCAAATMSX2TCAATTAATGGATCF-1AACAAAGMSX2TCAATTAATGGATCF-1AACAAAGMSX2TCAATTAATGGATCF-1AACAAAGNACT5TGCAAATTTCF-1TCAAGTGTANK32TCAATTAATGGATCF-1AACAAAGNK32TCAATTAATGGATCF-1TCAAGTGTANK32TCAATTAATGGATCF-1TCAAGTGTANK32TCAATTAATGGATCF-1TCAAGTGTANK32TCAATTAATGGATCF-1TCAAGTGTANK32TCAATTAATGGATCF-1TCAAGTGTANK41TAATTGCATCF-1TCAAGTGTANK32TCAAGTGGGTT <td>HOXIILI</td> <td>CGTTAATTGG</td> <td>PBX3</td> <td>т т с а т т с а т</td>	HOXIILI	CGT TAAT TGG	PBX3	т т с а т т с а т
NormalianNORMATINATIANTACITINORMATINATIANTACITINORMATINATIANTACITINORMATINATIANTACITINORMATINATIANTACITIPITXITAATCCPITX2TAATCCHOXB7TTATCPOUIFITAATCCPOUIFITAAACGACPOUSFITAAACGACPOUSFITAAACGACPOUSFIT AACGAAATPOUSFIT TAACGAAATTTAACGACPOU4FIC C C ATTAACGACPOUSFIA T G C A A A TT TAACGACPOU4FIC A A A TT TAACGACPOU4FIC A A A TT TAACGACPOU4FIA T G C A A A TT TAACGACPOU4FIA T G C A A A TT TAATTAPOU4FIA T G C A A A TT TAATTAPOU4FIA T G C A A A TT T T A T T APOU4FIA T G C A A A TT T A A T T AC C T C A T T A A T AIL2T T A A T T G CT T A A T T AC C C C A T T A A TMOX2T C A A T T A T T	HOX11L2	Ссттаатисс	PBX3-HOXC6	G C
HOAND TAATTGA PITAL PITAL TAATCC C HOXA9 TTTC PITX2 TAATCC C HOXB7 TTATGA CAC PIX1 TAATA HOXB8 TTTATGA C POUFI TAATA HOXB9 TTATC GAC POUFI TAAAT HOXC11 TTTAC POUFI TAAAT HOXC12 TTAACGA C POUSI TG CAAAT HOXC13 TTTAC GAC POUSI TG CAAAT HOXC14 TTTAC GAC POUSI TG CAAAT HOXC14 TTTAC GAC POUSI TG CAAAT HOXC15 TTAACGA C POUSI TG CAAAT HOXC14 TTTAC GAC POUSI GC TC ATTAACGAG HOXD3 TTTACGA G POUSI GC TC ATTAACGAG HOXD3 TTTACGA G C POUSI GC TC ATTAACGAG HOXD3 TTTACGA G C POUSI ATG CAAAT HOXC15 TTAATATCTG G G POUSI ATG CAAAT HOXC16 TTAATATC G G G POUSI ATG CAAAT HC16 TTTACGA G C T TAATATCTG G G POUSI ATG CAAAT HC16 TTTACGA G C T TAATATCTG G G POUSI ATG CAAAT HC16 TTTAAGTA POUFI GC TC ATTAAC C HC17 TTAAGTA POUFI GC TC ATTAAC C HC18 TG CAATTAATTG A POUFI ATG CAAAT HC19 TTAATTA G A POUFI ATG CAAAT HC19 TG CAATTAATTG A POUFI ATG CAAAT HC19 TG CAATTAATTG A POUFI ATG CAAAT HC19 TG CAAATTAATTG A POUFI ATG CAAAG HC19 TG CAAATTAATTG A POUFI ATG CAAGG CT G CT TAATG A AG HC19 TG CAAATTAATTG A POUFI ATG CAAAG HC1	HOVAIO	T	DITEXT	GIGAIIAIIACIII
HOXA9T T T T A T G A CPITX2T A T T C CHOXB7T T T A T G A CPMX1T A A T AHOXB8T T T A T G A CPNX1T G A C A GHOXB9T T A A C G A CPOUIFIT A A A THOXC11T T T A C G A CPOU2FIA T G C A A A THOXD12T T A A C G A CPOU3FIT G G A G A NA T A G A GHOXD13T T T A C G A GPOU3F2A T G C A A A TIFFIC T C T A A T G G GPOU4F1A T G C A A A TISL1G C T T A A T A C G G GPOU4F2A T G C A A A TIFI2T T A A G T APOU4F2A T G C A T T A A CIMX1T C A A T T A G T APOU3FIA T G C A A A TISL2T T A A T T APOU4F2A T G C A A A TIMX2T C A A T T A T T APOU3FIA T G C A A A TMOX2T C A A T T A T T G APOU3FIA T G C A A A TMOX2T C A A T T A T T G APOU3FIA T G C A A A TMOX2T C A A T T A T T G APOU3FIA T G C A A A TMOX2T C A A T T A T T G APOU3FIA T G C A A A TMOX2T C A A T T A T T G APOU3FIA T G C C A T T A A CMSX-1C T A A T T G APOU3FIA T G C A A A TMSX-2T C A A T T A T T G APOU3FIA T G C A A A TMSX-2T C A A T T A A T T G APOU3FIA T G C A A A TMSX-1C T A A T T A C T G A C A G C T T A A C A A A GA T G C A A A TMSX-2T C A A T T G AT C T A A T G C A A A TA T	HOAAIU	TATTGA	FIIAI	TAATCC
HOXB7 TTTATGAC PMX1 TATA HOXB8 TTTATGAC PMX1 TAATA HOXB9 TTTAC POUIFI TAATA HOXB9 TTTACGAC POUIFI TAAAT HOXC11 TTTAC POU2FI ATGCAAAT HOXD12 TTAACGAC POU3F1 TGGAGANAATAGAG HOXD13 TTTACGAG POU3F2 ATGCAAAT IFF1 CTCTAAATGGG POU3F4 ATGCAAAT ISL1 GCTTAATATCTG POU4F1 AC TGCAAAT ISL2 TTAAGTA POU4F2 GCTCATTAAC TGGACAGCT POU4F3 GCTCATTAAC A T ISL2 TTAATA HOXD2 TCAATTAA MC2 TCAATTAAT GA POU4F3 GCTCATTAAC MEII TGACAGCT POU3F1 TAGCAAAT MC2 TCAATTAAT GA POU6F1 TAATGAGCTGCXTAAT MC2 TCAATTAAT GA POU6F1 TAATGAGCTGCXTAAT MC3 TCCAATTAATTGA POU6F1 TAATGAGCTGCXTAAT MC3 TCCAATTAATTGA CUGF1 TAATGAGCTGCXTAAT MC4 TCCAATTAATTGA TCTG TGF TGACCAA	HOXA9	т ттт.С	PITX2	TAATCC
HOXBS T T T A T G A C PNXI T G A C A G HOXB9 T T A A C G A C POUIFI T T A A C G A C HOXC11 T T T A C G A C POU2FI A T G C A A A T HOXD12 T T T A C G A C POU3FI T G G A G A N A A T A G A G HOXD13 T T T A C G A G POU3FI T G C A A A T HOXD14 T T T A C G G G POU3FI T G C A A A T HOXD15 T T T A C G G G POU3FI T G C A A A T HOXD16 T T T A C G G G POU3FI T G C A A A T HOXD17 T T T A C G G G POU3FI T G C A A A T HOXD18 T T T A T C T G POU3FI T G C C A T T A A C HOXD19 T T T A C G G G POU3FI T G C C A T T A A C HI1 G C T C T T T A A C G G G POU4FI G C T C A T T A A C HI2 T T A A G G T A POU4FI G C T C A T T A A C HI1 T G A C A G C T POU4FI G C T C A T T A A C HI1 T G A C A G C T POU4FI G C T C A T T A A C HI1 T G A C A G C T POU3FI T A A T A HI1 T G A C A G C T T G A A T G C T C A T T A A C HI1 T G A C A G C T T G A A C A A A G HI1 </td <td>HOXB7</td> <td>TTATGAC</td> <td>PMX1</td> <td>TAATA</td>	HOXB7	TTATGAC	PMX1	TAATA
HOXB9 T T T T T T T HOXD12 T	HOXB8	TTATGAC	PNX1	TGACAG
HOXCII T T T T T T T T T T T T T T T T T T	HOXB9	T	POU1F1	
HOXEII T T T A C POU2FI A T G C A A A T HOXD12 T T A A C G A C POU3FI T G G A G A N A T A G A G HOXD13 T T T A C G A G POU3F2 A T G C A A A T IPF1 C T C T A A T G G G G POU3F4 A T G C A A A T ISL1 G C T T A A T A T C T G POU4F1 A T G C A A A T ISL2 T T A A G T A POU4F2 G C T C A T T A A C IH2 T T A A T A T C T G POU4F3 G C T C A T T A A C MEI1 G C T C A T T A A T T G A POU6F1 A T G C A A A T MSX-1 C C T A A T T G A POU6F1 A T G C A A A T MSX-2 T C A A T T A A T T G A POU4F3 G C T C A T T A A C MSX-2 T C A A T T A A T T G A POU6F1 A T G C A A A T MSX-2 T C A A T T A A T T G A POU6F1 A T A A T A MSX-2 T C A A T T A A T T G A POU6F1 A T A A T A MSX-2 T C A A T T A A T T G A POU6F1 A T A A T A MSX-2 T C A A T T A A T T G A T C A A A A T T G A C A A G N-0CT5 A T G C A A A T T G G C T G C X T A A T A T G A C A A G NK61 T T A A T A C T T C A A G T G C T G C T T T C A A G T G T T	UOYCU	T		TAAAT
HOXD12 T T T POU3FI T G	HUXCII	TTTAC	POU2FI	ATGCAAAT
HOXDI3 TTACGAG POU3F2 ATGCAAT IPFI CTCTAATGGG POU3F4 ATGCAAAT ISL-1 GCTTAATATCTG POU4F1 GCTCATTAAC TGGCTCATTAACCAAT ISL2 TTAATATCTG POU4F2 GCTCATTAAC IH2 TTAATTA MEII GCTCATTAATCC MEII GCTCATTAATCC MSX-1 CTAATTGA POU4F3 GCTCATTAAC MSX-2 TCAATTAATTGA POU4F3 GCTCATTAAC MSX-2 TCAATTGA POU4F3 GCTCATTAAC MSX-2 TCAATTAATTGA POU4F3 GCTCATTAATGAGCTGCXTAAT MSX-2 TCAATTAATTGA POU4F3 GCTCATTAATGAGCTGCXTAAT MSX-2 TCAATTAATTGA POU4F3 GCTCATTAATGAGCTGCXTAAT MSX-2 TCAATTAATTGA POU4F3 GCTCATTAATGAGCT MSX-2 TCAATTAATTGAA TCF-1 AACCAAAG N-OCT5 ATTGAATTGA TCF-1 TCAAGTGTT NK61 TTAATTAC TTF-1 TCAAGTGTT	HOXD12	T	POU3F1	
IPF1 T T R R R ISL1 G C T T A A T A T C T G POU4F1 G C T C A T T A A C ISL2 T G G POU4F2 A T ISL2 T T A A G T A POU4F2 G C T C A T T A A C IH2 T T A A G T A POU4F3 G C T C A T T A A C MEI1 G C A C C A T T A A C A T MX2 T C A A T T A A T T G A POU6F1 T A A T G A A T MSX-1 C T C A A T T A A T T G A POU6F1 T A A T A MSX-2 T C A A T T A A T T G A TCF-1 A A C A A A G NK61 T T A A T T A C TTF-1 T C A A G NK61 T T A A T T A C TTF-1 T C A A G G C T C	HOXD13	TTTACGAG	POU3F2	TGGAGANAATAGAG ATGCAAAT
ISL ¹ C T C T A A T G G G POU4F1 G C T C A A A T ISL ¹ G C T T A A T A T C T G POU4F1 G C T C A T T A A C T G G A T ISL ² T T A A G T A POU4F2 G C T C A T T A A C HH2 T T A A T T A POU4F3 G C T C A T T A A C MEII G A C A G C T MOX2 T C A A T T A A T T G A POU5FI A T G C A A A T MOX2 T C A A T T A A T T G A POU6FI T A A T G A G C T G C X T A A T MSX-1 C T A A T T A A T T G A POU6FI C A T A A T A MSX-2 T C A A T T A A T T G A TCF-1 A A C A A G N-OCT5 A T G A A A T TGF T G A C A A T NK61 T T A A T A C TTF-1 T C A A C A A G N-OX2 T C A A T T A A T T G C A A A T NK2.2 T C A A T T A A T T G C A A A T NK2.2 T C A A T T A A T G G C T C T C C T C C C C C C C C C C	IPF1	т	POU3F4	
ISL-1 GCTTAATATCTG POU4F1 GCTCATTAAC TGGG GCTCATTAAC A T GCTCATTAC C GCTCATTAC GCTCATTAC C IH2 TTAATTA GC G OU4F3 GCTCATTAC G G OU4F3 GCTCATTAC G G OU5F1 ATGCAAAT MOX2 TCAATTAATTGA POU6F1 TAATGAGCTGCXTAAT C MSX-1 CTAATTGGA SATB1 C ATAATA MSX-2 TCAATTAATTGA TCF-1 AAC C NK2.2 TCAAGTGGTT		CTC TAAT GGG	100011	ATGCAAAT
ISL2 TRAGTA POU4F2 GCTCATTACC GCTCATTACC LH2 GCTCATTAC MEII GCCCATTAC MX2 TCAATTATTG POU5FI TGACAGCT POU5FI TGACAGCT POU5FI TGACAGCT CATTAC MX2 TCAATTAATTGA CCTGCXTAAT MX2 TCAATTAATTGA CCTGCXTAAT MX2 TCAATTAATTGA CC MX2 TCAATTAATTGA CC MX2 TCAATTAATTGA CC MX2 TCAATTAATTGA CC TGACAA C NOCTS TCAATTAATTGA CC TGACAA C NKC1 TTAATTAC TIGIF TGACAA C TCAAGTGCT TCAAGTGGTGCT	ISL-I	GCT TAAT ATCTG	POU4F1	A T GCTC ATTA AC
TTAAGTA ICOM GCTCATTAAC LH2 TTAATTA POU4F3 GCTCATTAAC MEII G GCTCATTAAC MX2 TCAATTAATTGA POU5FI ATGCAAAT MX2 TCAATTAATTGA POU6FI TAATGAGCTGCXTAAT MSX-1 CTAATTGG SATBI C NSX-2 TCAATTAATTGA TCF-1 AACAAAG NK61 TTAATTAC TTF-1 TCAAGTGTT NK2.2 TCAAGTGGTT TCAAGTGTT	ISL-2	T GG	POU4F2	А Т
LH2 TTAATTA POU4F3 G C T C ATTAAC MEI1 G G C T C ATTAAC MX2 T C A A T T A ATT G A POU6F1 ATG C A A T MSX-1 C T A ATT G A POU6F1 T A AT G A G C T G C X T A A T MSX-2 T C A A T T A ATT G G A TCF-1 A C C A A A G N-OCTS A T G C A A A T TGIF T G A C A A NK61 T T A ATT A C TTF-1 T C A A G T G T T NK2.2 T C A A G T G G T T T C A AG T G G T T		T T A A G T A	100112	GCTCATTAAC
MEI1 G POUSF1 ATGCAAAT MOX2 T C A A T T A T G A POU6F1 T A A T G A G C T G C X T A A T MSX-1 C T A A T T G A SATB1 C MSX-2 T C A A T T A T T G G A TCF-1 A T A A T A N-OCT5 A T G C A A A T TGIF T G A C A A NK61 T T A A T T A C TTF-1 T C A A G T G T T NK2.2 T C C A G T G C T T T C A A G T G T T	LH2	т таат та	POU4F3	GCTC ATTA AC
TGACAGCT NOSTI ATGCAAAT MOX2 TCAATTAATTGA POU6FI TAATGAGCTGCXTAAT MSX-1 CTAATTG SATBI C MSX-2 TCAATTAATTGA TCF-1 AACAAAG N-OCT5 ATGCAAAT TGIF TGACAA NK61 TTAATTAC TTF-1 TCAAGTGTT NKX2.2 TCAAGTGGTT TCAAGTGGTT	MEII	G	POUSEI	
MOX2 TCAATTAATTGA POUDFI TAATGAGCTGCXTAAT MSX-1 CTAATTG SATBI C MSX-2 TCAATTAATTGA TCF-1 AACAAAG N-OCT5 ATGCAAAT TGIF TGACAA NK61 TTAATTAC TTF-1 TCAAGTGTT NKX2.2 TCAAGTGGTT AC	MOV2	TGACAGCT	POLICE	ATGCAAAT
MSX-1 C T A A T T G SATB1 A T A A T A MSX-2 T C A A T T A A T T G A T CF-1 A A C A A A G N-OCT5 A T G C A A A T T GIF T G A C A A NK61 T T A A T T A C T TF-1 T C A A G T G G T T NKX2.2 T C A A G T G G G T T T C A A G T G G G T T	MOA2	T C A A T T A A T T G A	rouori	TAATGAGCTGCXTAAT C
MSX-2 T C A A T T A A T T G A T CF-1 A A C A A A G N-0CT5 A T G C A A A T T GIF T G A C A A NK61 T T A A T T A C T TF-1 T C A A G T G T T NK2.2 T C A A G T G G G T T T C A A G T G G G T T	MSX-1	CTAATTG	SATB1	а таат а
NOCIS ATGCARAT TGIF TGACAR NK61 TTAATTAC TTF-1 TCAAGTGTT NK2.2 T AC	MSX-2	Τ C A A T T A A T T G A	TCF-1	AACAAG
T AC TCAAGTGTT NKX2.2 TCAAGTGGTT	N-OCT5 NK61	· ATGCAAAT	TGIF	TGACAA
NKL2.4 T C A A G T G G T T	NICO1	T AC	117-1	TCAAGTGTT
	INKAZ.2	T CAAG TGGTT		

Figure 1. DNA binding sites for human homeodomain loci implicated in genetic or genomic disorders. Data is derived from the published literature, citations in OMIM, and entries for DNA-bound homeodomain structures from the Protein Data Bank. The core regions of each of the DNA binding sites are shown in bold type. An expanded version of this table is available online (at http://genome.nhgri.nih.gov/homeodomain); the expanded version includes a listing of alternative gene names, as well as citation information.

A new feature of the Homeodomain Resource is the inclusion of DNA binding sites for human homeodomain loci implicated in genetic or genomic disorders (Fig. 1). These data were obtained by extensive review of the published literature, citations in OMIM, and entries for DNA-bound homeodomain structures from the Protein Data Bank. The online version of Figure 1 (available via http://genome.nhgri.nih.gov/homeodomain) includes alternate gene names and references to the primary citation from which the information was retrieved.

REFERENCES

 Gehring, W., Qian, Y., Billeter, M., Furukubo-Tokunaga, K., Schier, A., Resendez-Perez, D., Affolter, M., Otting, G. and Wuthrich, K. (1994) Homeodomain-DNA recognition. *Cell*, 78, 211–223.

- 2. Laughon, A. (1991) DNA binding specificity of homeodomains. *Biochemistry*, **30**, 11357–11367.
- Dekker, N., Cox, M., Boelens, R., Verrijzer, C., van der Vliet, P. and Kaptein, R. (1993) Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature*, 362, 852–855.
- 4. Viglino, P., Fogolari, F., Formisano, S., Bortolotti, N., Damante, G., Di Lauro, R. and Esposito, G. (1993) Structural study of rat thyroid transcription factor 1 homeodomain (TTF-1HD) by nuclear magnetic resonance. *FEBS Lett.*, **336**, 397–402.
- Gruschus, J., Tsao, D., Wang, L., Nirenberg, M. and Ferretti, J. (1997) Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. *Biochemistry*, 36, 5372–5380.
- Kissinger, C., Liu, B., Martin-Blanco, E., Kornberg, T. and Pabo, C. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell*, 63, 579-590.
- Ceska, T., Lamers, M., Monaci, P., Nicosia, A., Cortese, R. and Suck, D. (1993) The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB1/HNF1 and implications for DNA binding. *EMBO J.*, **12**, 1805–1810.

- Liu,B., Kissinger,C. and Pabo,C. (1990) Crystallization and preliminary X-ray diffraction studies of the engrailed homeodomain and of an engrailed homeodomain/DNA complex. *Biochem. Biophys. Res. Commun.*, **171**, 257–259.
- Qian,Y., Billeter,M., Otting,G., Muller,M., Gehring,W. and Wuthrich,K. (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, 59, 573–580.
- Qian, Y., Furukubo-Tokunaga, K., Resendez-Perez, D., Muller, M., Gehring, W. and Wuthrich, K. (1994) Nuclear magnetic resonance solution structure of the fushi tarazu homeodomain from Drosophila and comparison with the Antennapedia homeodomain. J. Mol. Biol., 238, 333–345.
- Wolberger, C., Vershon, A., Liu, B., Johnson, A. and Pabo, C. (1991) Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell*, 67, 517–528.
- Buerglin, T. (1994) In Duboule, D. (ed.), *Guidebook to the Homeobox Genes*. Oxford University Press, Oxford, UK, pp. 25–71.
- Banerjee-Basu,S., Ryan,J.F. and Baxevanis,A. (2000) The Homeodomain Resource: sequences, structures and genomic information. *Nucleic Acids Res.*, 28, 329–330.