

A 7.5 Mb Sequence-Ready PAC Contig and Gene Expression Map of Human Chromosome 11p13–p14.1

Beate Gawin,¹ Alexandra Niederführ,¹ Nina Schumacher,¹ Holger Hummerich,² Peter F.R. Little,² and Manfred Gessler¹

¹Physiologische Chemie I, Biozentrum der Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany; ²Department of Biochemistry, Imperial College of Science, Medicine and Technology, London SW7 2AZ

The region p13 of the short arm of human chromosome 11 has been studied intensely during the search for genes involved in the etiology of the Wilms' tumor, aniridia, genitourinary abnormalities, mental retardation (WAGR) syndrome, and related conditions. The gene map for this region is far from being complete, however, strengthening the need for additional gene identification efforts. We describe the extension of an existing contig map with PI-derived artificial chromosomes (PACs) to cover 7.5 Mb of 11p13–14.1. The extended sequence-ready contig was established by end probe walking and fingerprinting and consists of 201 PAC clones. Utilizing bins defined by overlapping PACs, we generated a detailed gene map containing 20 genes as well as 22 anonymous ESTs which have been identified by searching the RH databases. RH maps and our established gene map show global correlation, but the limits of resolution of the current RH panels are evident at this scale. Initial expression studies on the novel genes have been performed by Northern blot analyses. To extend these expression profiles, corresponding mouse cDNA clones were identified by database search and employed for Northern blot analyses and RNA in situ hybridizations to mouse embryo sections. Genomic sequencing of clones along a minimal tiling path through the contig is currently under way and will facilitate these expression studies by in silico gene identification approaches.

One step towards the functional understanding of the human genome is large-scale sequencing followed by sequence-based gene identification. Sequencing is performed in a directed strategy to reduce redundancy to a minimum, as applied to the *Caenorhabditis elegans* genome (*C. elegans* consortium 1998). This strategy relies on prior mapping information and sequence-ready contigs as resources for the generation of a minimal tiling path to be followed during sequencing. More recently, a random sampling strategy, as applied in bacterial sequencing, has been suggested to also be practicable for the human genome. To assess the potential of this strategy, the *Drosophila melanogaster* genome is being used as a model (Pennisi 1999b). Sequencing of cDNAs from a large variety of different libraries (i.e., IMAGE consortium; Lennon et al. 1996) has been employed as a shortcut for gene identification. However, the neglect of mapping information, as well as information hidden in intronic sequences or promoters with their regulatory elements, is a drawback of this method. Moreover, genes with very low or highly restricted expression still pose a severe problem as they are underrepresented in the commonly used cDNA libraries. Nevertheless, partial information on a signifi-

cant fraction of genes and their corresponding gene products can be obtained in less time and at lower cost compared to genomic sequencing (Adams et al. 1991; Hillier et al. 1996). With the recent proposal to speed up genome sequencing (Pennisi 1999a; Wadman 1999), these ESTs can soon be integrated into the emerging human genome sequences.

We have focused on the genetic content of the region p13 on human chromosome 11. Intense analysis of the deletion region of patients with WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities and mental retardation) has led to the isolation of the *WT1* and *PAX6* genes (Call et al. 1990; Gessler et al. 1990; Ton et al. 1991). However, the molecular mechanisms underlying their regulation are still unclear. For *PAX6*, position effects detected are caused by rearrangements well downstream of the known exons (Fantès et al. 1995; Schedl et al. 1996). The tissue-restricted expression of *WT1* appears to be driven by far away *cis*-regulatory elements that have yet to be identified (Scholz et al. 1997). The knowledge of the complete genomic sequence flanking these two genes and subsequent interspecies comparison with corresponding mouse or fugu data will be an essential step towards the elucidation of these mechanisms (Miles et al. 1998). Mental retardation and growth retardation, which is seen in ~30% of WAGR patients (Turleau et al. 1984), represent two other features of

¹Corresponding author.
E-MAIL gessler@biozentrum.uni-wuerzburg.de; FAX 49-931-888-4150.

this syndrome but no candidate genes have yet been identified. Frequent allele loss in distinct regions of 11p13 suggests the presence of one or more tumor suppressor genes other than *WT1* that may be involved in a large variety of tumors, including lung, bladder, breast, and ovarian cancers (Bepler and Garcia-Blanco 1994; Kiechle-Schwarz et al. 1994; Shipman et al. 1998; Shipman et al. 1993; Vandamme et al. 1992).

In order to create the tools for sequence based gene identification, we extended a 3 Mb sequence-ready PAC contig described previously by Niederführ et al. (1998) to cover a region of 7.5 Mb of 11p13 reaching into p14.1. We also report the characterization of 22 novel 11p13 specific EST clones, their integration into the physical map, and their initial expression analyses.

RESULTS

Contig Construction

Niederführ et al. (1998) have constructed a 3 Mb sequence-ready PAC contig that covers the inner part of 11p13. We have extended this contig in the telomeric (*t*; PAX6-D11S155) and centromeric (*c*; D11S935-CAT) direction by hybridizing high density grid filters of the RPCI-1, -3, -4, and -5 PAC libraries with 11p13 specific markers and probes derived from a YAC contig covering 11p13 (Gawin et al. 1995). 125 PAC clones for the centromeric region *c* and 94 for the telomeric region *t* were identified and used as starting material for further end probe walking.

The mapping strategy followed Niederführ et al. (1998). End probes of each PAC were generated by linker mediated PCR and hybridized to colony filters of the arrayed sublibrary to identify overlapping clones. Clone insert sizes were determined by *NotI* restriction digests of the respective PAC clones. Additionally, we used markers that were already integrated into a detailed physical map of 11p13 as hybridization probes (Gawin et al. 1995). The combination of these data sets allowed an accurate alignment of the analyzed PAC clones along the physical map of 11p13.

With the first set of 219 PAC clones we could establish 10 individual, preliminary contigs (*c*:4; *t*:6). To close intercontig gaps between the new clone assemblies and to link them to the *CAT-PAX6* contig, the RPCI PAC libraries were rescreened with end probes of clones located at the outer edges of these contigs. Two more cycles of clone ordering and rescreening led to the identification of 199 additional clones (*c*:43; *t*:156). Out of a total of 418 clones, we employed 125 (*c*:36; *t*:89) to build the two final contigs. Clones with either repetitive sequences at the insert ends that inhibited successful hybridizations, or with poor growth on colony filters were rejected from further analysis. Similarly clones that turned out to be located outside our region of interest were disregarded. Another 61 PAC

clones that were part of the working draft contigs, are not included in the final version due to abundant coverage of several regions or failure to determine insert sizes by PFGE, likely due to large inserts prone to instability. The ordered clone contig is shown in Figure 1.

A complementary approach for contig construction was followed by performing fingerprint analyses on all PAC clones that had been identified during the screening process. The comparison of *HindIII/Sau3AI* digest patterns of all clones strongly corroborates the overlaps suggested by the hybridization data without apparent inconsistencies. However, in the few regions with single clone coverage or short overlaps, fingerprinting could not reach statistical significance (a random probability of $< 10^{-4}$, data not shown). In these cases, overlap is supported by a minimum of three shared hybridization markers. Only two reciprocal hybridizations were detected with two overlapping PACs (815A16/562D20) centromeric to the *FSHB* locus, pointing to an underrepresentation of this genomic DNA fragment in the libraries used. All fingerprinting data are stored in the chromosome 11 database 11DB (<http://chr11.bc.ic.ac.uk>).

Combined with the previously described 3 Mb *CAT-PAX6* contig of Niederführ et al. (1998), we have now constructed a sequence-ready PAC contig covering approximately 7.5 Mb of 11p13-14.1. This contig extends from the marker D11S935 at the centromeric end to 250 kb beyond the marker D11S155 located in band p14.1. Our newly established contig extensions comprise 36 and 89 PAC clones for the centromeric and telomeric regions respectively, with an average 3.75-fold coverage.

Identification of 11p13 Specific Transcripts-Gene Map

The complete PAC contig represents an excellent tool for the positioning of genes. The nested nature of the PAC clones and their exact anchoring along the physical map, allows the localization of genes within an interval of generally < 100 kb with respect to known markers.

All available versions of radiation hybrid (RH) maps were searched for ESTs assigned to 11p13 (especially <http://ncbi.nlm.nih.gov/genemap98>; <http://shgc.stanford.edu/Mapping/rh/MapsV2/index.html>; http://carbon.wi.mit.edu:8000/cgi-bin/contig/phys_map). We concentrated mainly on three GeneMap'98 RH intervals defined by the markers D11S1324 (*t*) and D11S4102 (*c*) that flank the PAC contig. To compensate for possible differences between the statistically based RH maps and our physical data, we also included the region D11S1324-S904 telomeric to *FSHB*, and a subset of clones from RH bins located centromeric to *RAG1/2* in 11p12 (Table 1).

With the identified RH map entries we performed extensive similarity searches and UniGene cluster com-

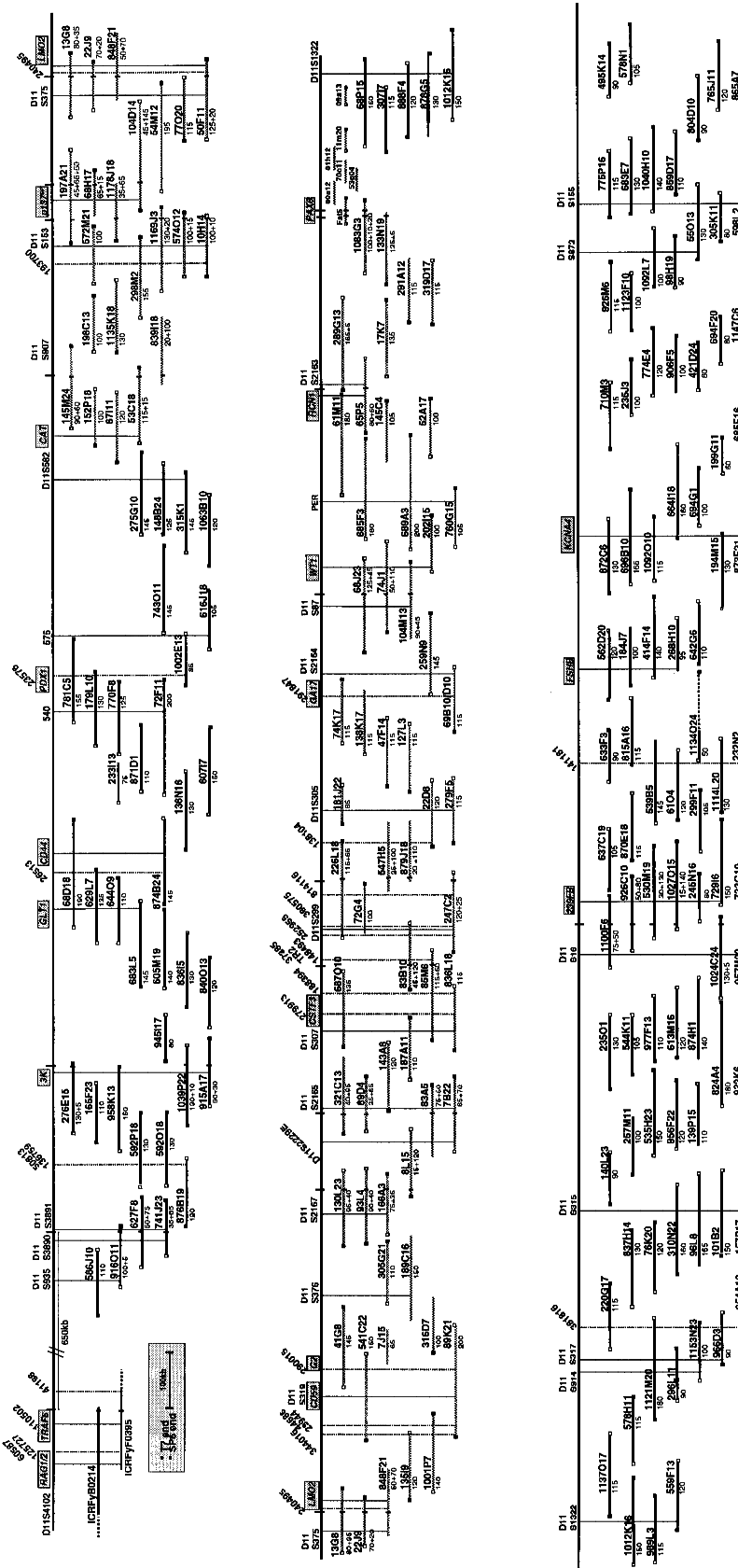


Table 1. Listing of Putative 11p13–14.1-Specific Transcripts

GeneMap'98 entries	GeneMap'98 intervals	Unigene cluster RH-marker	Representative cDNA clones	Mapped to 11p13	Northern hybridization
WI-9523	D11S929–D11S1324	Hs.91791	279594	–	n.d.
stSG8940		Hs.79793	148452	–	n.d.
WI-15111		Hs.32478	52631	–	n.d.
SHGC-15981		Hs.119020	452395	–	n.d.
A006R28	D11S904–D11S1324	RH25969	75955	–	n.d.
stSG12676		RH36159	191848	–	n.d.
A006U38		Hs.91393	31804	–	n.d.
stSG4282		Hs.119631	123100	–	n.d.
stSG35677/A42SGC33830	D11S904–D11S914	RH51732	121149	–	n.d.
AA251905		Hs.48473	360671	–	n.d.
WI-13781		Hs.23165	380575	+	+
WI-16368	D11S1324–D11S914	Hs.148136	252687	–	n.d.
WI-13900		Hs.24997	37505	–	n.d.
WI-17439		Hs.102530	141181	+	+
WI-17370		Hs.146441	138104	+	+
WI-9460		Hs.99014	148493	+	+
sts-AA017167		Hs.32713	381816	+	–
H44825		Hs.33130	188394	+	–
sts-N32907	D11S914–D11S907	Hs.44402	259883	–	n.d.
D11S2259E (G2)		Hs.22270	290015	+	+
D11S2229E		Hs.153900	31A10	+	+
R51067		Hs.22383	37385	+	+
sts-H37799		Hs.32683	244989	–	n.d.
sts-N37061		Hs.43561	273773	–	n.d.
A004V27		Hs.99766	44686	+	+
sts-R49201		Hs.141002	37144	–	n.d.
sts-H90759		Hs.128746	240495	+	+
SGC44450		Hs.103165	344016	+	–
WI-11454		Hs.16577	29944	+	+
stSG9634		Hs.22666	252955	+	–
WI-15813		Hs.71472	193700	+	+
A003A07 (PDX1)		Hs.74642	23576	+	+
WI-12839 (CSTF3)	D11S907–D11S4102	Hs.82642	279913	+	+
D11S2280E/2323E		Hs.4774	26513	+	+
WI-20739		Hs.181766	25156	–	n.d.
WI-12351/D11S981E		Hs.14512	136759	+	+
SGC32165/stSG25954		Hs.16944	50813	+	+
stSG15531		Hs.28723	140811	–	n.d.
n.a.		Hs.8253	60587	*	n.d.
SGC34020		Hs.19987	125727	*	n.d.
overlapping WI-17696 (TRAF6)		Hs.31171	110502	*	n.d.
WI-21989		Hs.26172	588550	–	n.d.
stSG-3901		Hs.26505	41188	*	+
A006H38/SHGC-31880	D11S4102–D11S1355	Hs.106863	23368	–	n.d.
WI-6923	D11S1330–D11S1355	Hs.106826	254585	–	n.d.
SGC31982/A006G11	D11S1330–D11S1361	Hs.9848	229775	–	n.d.
stSG31410	D11S1355–D11S1361	Hs.71917	624651	–	n.d.
stSG3202		Hs.85889	32850	–	n.d.
n.a. (GA17)	DXS983–DXS995	Hs.69469	291847	+	+
n.a. (TR2)			TR2	+	+
n.a. (FJX1)			NE3	+	+
n.a.		NE70/D11S3899	814116	+	–

Human 11p13–14.1 IMAGE cDNA clones. Clones marked by “n.a.” (not annotated) are not derived from the current GeneMap'98 but as described in the text. + and – in the mapping column indicate clones that are or are not mapped to the PAC contig, * marks clones that have been mapped to 11p13 by employing YAC clones from a contig that extends further centromeric. + and – in the Northern analysis column point to the presence or absence of hybridization signals, hybridizations marked by n.d. (not done) have not been performed because the probe did not map to the PAC contig.

parison to eliminate redundancy arising through multiple entries of sequences that are likely derived from the same transcript. Several ESTs could be assigned to 14 genes mapped previously to our region of interest (*cen-RAG1/2*, *GLT1/SLC1A2*, *CD44*, *CAT*, *p137^{GPI}/M11S1*, *LMO2*, *CD59*, *G2*, *WT1*, *RCN1*, *PAX6*, *FSHB*, *KCNA4-tel*). We identified an additional 47 nonredundant EST clusters or single ESTs and selected one representative cDNA clone from each for further analyses (Table 1). Three of the 47 ESTs correspond to known genes, namely the *tumor necrosis factor receptor-associated factor 6* (*TRAF6*), the X component of the pyruvate dehydrogenase complex (*DLAT/PDX1/E3BP*) and the 77 kD subunit of the *cleavage stimulation factor* (*CSTF3*). Whereas *CSTF3* and *TRAF6* had not yet been assigned to 11p13, *PDX1* has been mapped previously to 11p13 by FISH analysis (Aral et al. 1997; Ling et al. 1998). We also included four cDNAs from sources other than the RH maps: The cDNA sequence of the GA17 protein (GenBank accession no. AF064603) shows similarity to the Fugu cosmid clone 38e12, neighboring the Fugu *WT1*, *RCN1* and *PAX6* gene loci, suggesting a possible synteny (Miles et al. 1998; C. Miles and G. Elgar, pers. comm.). The fact that *GA17* maps to chromosome X (DXS983–995) according to GeneMap'98 likely results from the presence of a truncated *GA17* pseudogene that can be detected in the PAC 849L7 located on Xq21 (GenBank accession no. AL008987). The similarity of the 1.25 kb cDNA sequence of *GA17* (GenBank accession no. AF060643) to this PAC clone is restricted to the last 300 bp of the 3' end without any indication for splice sites. By using the STS A002N23 located within these 300 nucleotides for mapping *GA17* on the RH map, the intronless X-derived sequence is being amplified preferentially, explaining the observed mapping inconsistency. The *TR2* and *3K* cDNAs (GenBank accession nos. AJ245600 and AJ245599, respectively), the latter has been identified as the human homolog *FJX1* of the *D. melanogaster fj* gene (Ashery-Padan et al. 1999), have previously been identified by CpG island analyses on 11p13 and subsequent cDNA selection (Thäte et al. 1995). The IMAGE clone 814116 shows sequence similarity to another CpG island clone NE70 (D11S3899) located on 11p13.

To determine the location of all novel cDNA clones within the PAC contig, we hybridized them to colony filters of the 11p13-specific PAC sublibrary. Out of 52 independent clones tested, 28 could finally be assigned to our region of interest (Fig. 1). Four of these 28 clones were located outside of the PAC contig to the centromeric border of 11p13, as they hybridized to YAC clones from a contig that extends further centromeric (Gawin et al. 1995).

The integration of all hybridization data into the existing physical map resulted in the generation of a

detailed gene map for the region 11p13–14.1 containing 20 genes, as well as 22 anonymous transcripts (Fig. 1).

Comparison Between RH and PAC Contig Mapping Data

Current RH maps represent a rich source of detailed mapping information that is readily available to the research community. To assess the concordance of the data obtained by this statistical method and our physical approach, we graphically correlated the localization of GeneMap'98 entries within the D11S4102–D11S1324 interval with the positions of corresponding markers on the contig map as shown in Figure 2.

In cases where one transcript is characterized by multiple independent EST-derived entries in GeneMap'98, only one representative cDNA clone was employed for hybridization to the PAC contig (Table 2). In one instance, three seemingly independent markers may actually be derived from a single transcription unit as suggested by the alignment of cDNA sequences shown in Figure 3.

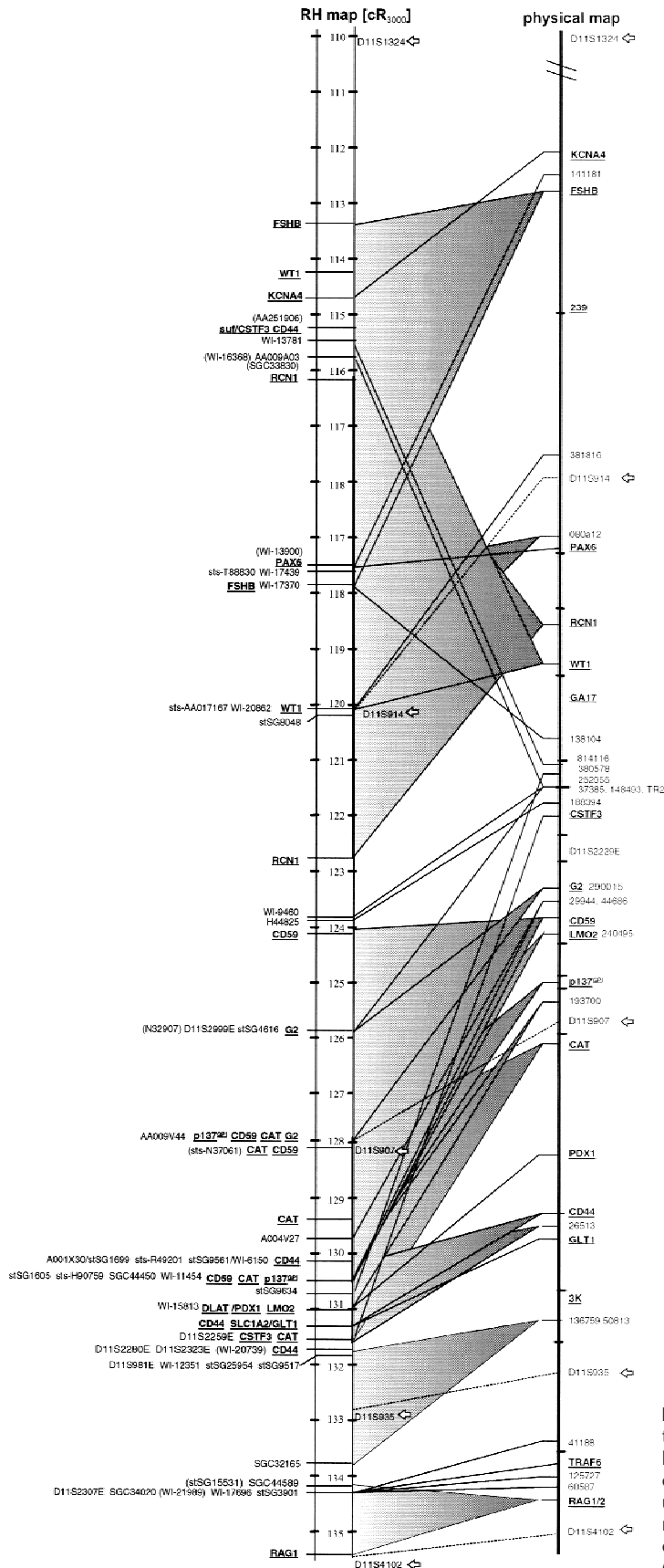
Generally good overall correspondence between the two data sets can be observed with local variations for several transcripts of up to 6 cR generated by multiple RH entries. In three cases (*CD44*, *CSTF3*, and cDNA contig 37385) these mapping differences were much larger, extending to 10–15 cR (compare Table 2). The estimate that 1 cR₃₀₀₀ is equivalent to ~270 kb on chromosome 11 (Gyapay et al. 1996) is in agreement with our own findings for 11p13, where 1 cR₃₀₀₀ corresponds to ~300 kb. However, out of 34 clones listed in the RH interval D11S1324–D11S4102 that is completely covered by our PAC contig, only 24 (71%) do map to the corresponding contig region. This fact, together with observed mapping uncertainties regarding order and distance between single entries, points to the difficulties when translating statistically derived coordinates into basepairs.

Human Northern Blot Analyses

To generate expression profiles for all new transcripts assigned to our PAC contig, we hybridized the corresponding human IMAGE clone inserts to Northern blots containing RNA from multiple human tissues as summarized in Figure 4.

Out of a total of 25 cDNA clones employed, we could detect signals on the Northern blots with 20 clones. Failure to detect transcripts for the other cDNA clones might be due to low expression or expression in tissues other than those employed on the Northern blots. The fact that some of the EST clones may be derived from contaminating genomic DNA may also play a role.

In general, rather ubiquitous expression of the analyzed transcripts was detected. However, some are obviously restricted to certain tissue types like liver



(TR2) or brain (148493, 290015, and 26513). In two cases (380575 and 290015), two bands on the autoradiogram point to alternative splice or polyadenylation variants of the respective transcripts. The similarity in size and expression pattern of the transcript detected by 290015, as well as the genomic localization, provide strong evidence that this cDNA may be part of the G2 transcript (Glaser et al. 1989). However, it does not match the published 5 kb partial cDNA sequence (GenBank accession no. U10991).

In three cases, cDNA clones from different EST entries colocalized on the PAC contig (50813/136759, 44686/29944, 148493/37385/TR2), but extensive BLAST search with additional overlapping sequences did not reveal any connection between them. The similarity of expression patterns and transcript sizes suggests that 50813 and 136759 are nevertheless derived from the same gene, while 44686/29944 and 148493/37385/TR2 do indeed represent independent transcription units each (Fig. 4).

Expression Analyses with Mouse cDNA Clones

For those human cDNA clones that mapped to our PAC contig, we searched the mouse EST database for homologous clones. To render the similarity search more successful, we not only used the 3' and 5' end sequences of the respective IMAGE clones, but also where available, the annotated cDNA sequences of corresponding genes (3K/FJX1, PDX1, G2, TR2, and GA17). The murine pBs13 cDNA was also used to search for EST clones, as human clone 37385 is annotated as "highly similar to testis-specific mouse protein pBs13" in GeneMap'98, showing up to 42 % identity and 62 % similarity for this protein (summary Table 3).

Northern blots containing RNA of whole embryos and of various tissues of adult mice were probed with these mouse cDNA clones and results are summarized in Figure 4. The mRNAs of most transcripts are distributed ubiquitously in all tissues analyzed. Exceptions are 407691, 658218, ffx1, where a brain-specific expression in adult tissues is detected,

Figure 2 Comparison of GeneMap'98 (left) and PAC contig mapping data (right). Corresponding loci are connected by lines. GeneMap'98 entries that could not be mapped to our PAC contig are enclosed by brackets; gene names are underlined. Dotted lines and arrows mark RH framework markers. Shaded triangles describe the positional variation of genes or markers resulting from the use of multiple STSs (compare Table 2).

Table 2. Redundancy of GeneMap'98 Entries

Genes/EST clusters	GenMap'98 entries (localization in cR ₃₀₀₀)						Hybridization probes
<i>RAG1</i>	SGC44589	134.29	<i>RAG1</i>	135.53	sts-M29474	135.53	s.c.
<i>CD44</i>	stSG35398	115.33	stSG9538	130.14	stSG12711	130.19	s.c.
	R36812	130.19	H96810	131.37	<i>CD44</i>	131.68	
	sts-N23307	131.68					
<i>CAT</i>	sts-H15685	128.00	sts-X04076	128.10	WI-19159	129.39	s.c.
	stSG41761	130.60	stSG21792	131.60			
<i>p137^{GPI}</i>	sts-T10644	128.00	stSG3783	130.60			s.c.
<i>CD59</i>	<i>CD59</i>	124.09	sts-M34671	128.00	WI-7543	128.10	s.c.
	stSG15292	130.60					
<i>G2</i>	SGC30174	125.87	U10991	128.00			290015
<i>CSTF3</i>	WI-12839	131.57	stSG30131	115.33			110502
<i>WT1</i>	<i>WT1</i>	114.28	M60614	120.18	X51630	120.18	s.c.
<i>RCN1</i>	stSG8897	116.08	GC33196	122.79			s.c.
<i>FSHB</i>	<i>FSHB</i>	113.44	SGC355392	117.83			s.c.
Hs.97925	stSG8048	120.28	sts-T88830	117.65	WI-20862	120.18	(80a12)
Hs.116577	WI-11454	130.60	A007J20	124.56			29944
IMAGE 50813	stSG25954	131.80	SGC32165	133.77			50813
cDNA contig 37385	A009A03	115.87	R51067	125.97	stSG15791	125.97	37385

Summary of multiple RHdb entries for human genes or EST clusters. The localization of the RH entries is indicated in cR₃₀₀₀. Hs. Human sequence clusters entered in the UniGene database. For each gene or EST cluster one representative hybridization probe was chosen to determine its corresponding position on the PAC contig map. For most of the known genes single copy probes (s.c.) have been used that are described in Gawin et al. (1995). For *G2*, *CSTF3* and the EST clusters, inserts of IMAGE clones have been employed. The human sequence cluster Hs.97925 that shows no similarity to *PAX6* is represented by the sequenced cosmid clone 80a12 located at the *PAX6* locus (see Figs. 1 and 2).

and 989128, which showed strong expression in liver and relatively weak expression in kidney. The IMAGE clone 603000 did detect a transcript in adult testis tissue only. However, size differences between transcripts of 37385 (human)/603000 (mouse) or 26513 (human)/676616 (mouse) provide strong evidence that the mouse cDNA clones identified do not represent the orthologous transcripts. For two of 14 independent clones analyzed, no corresponding mRNAs could be detected in the tissues used for the Northern blots (962716 and 670550).

Expression analyses by Northern blotting is restricted to whole mouse embryos or tissues of adult mice that can be obtained in sufficient quantity. Therefore, we performed RNA in situ hybridization on paraffin sections of E14.5 to E17.5 mouse embryos to detail the fetal expression data as shown in Figure 5.

The expression of the *G2* transcript detected in human, as well as in murine adult neuronal tissues, is

already present in the embryonic stage. The mouse embryo sections analyzed show a global but specific staining of neuronal tissues with an emphasis on the neopallial cortex along with expression in the retina, the developing teeth, and the thymus gland (Fig. 5A–C).

A very similar pattern but at lower levels can be observed for the anonymous IMAGE clone 658218 with the exception of a less explicit staining of the neopallial cortex region (Fig. 5D–F).

For *PDX1* we detected strong expression in skeletal muscle and heart which agrees with data obtained from the literature (Aral et al. 1997; Ling et al. 1998) as well as the preceding Northern blot analyses. Weaker expression of *PDX1* mRNA is also present in neuronal tissues as well as in the primordia of the teeth (Fig. 5G–I).

A high concentration of the *GAI7* mRNA is found in the mouse embryo in various organs, mainly lung, thymus, or the developing kidney, but also at lower

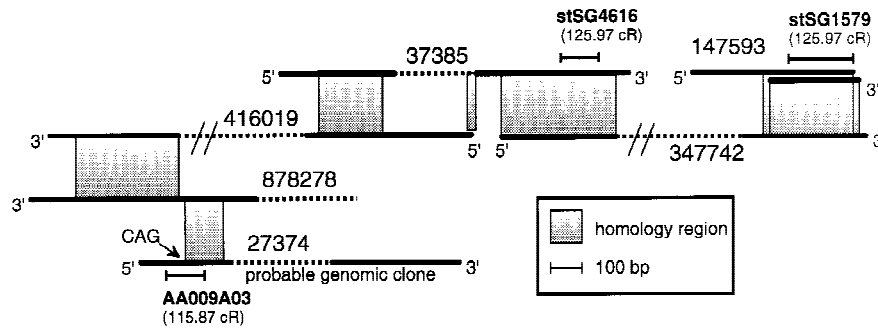


Figure 3 cDNA contig 37385 as established by BLAST search. Insert sizes of 27374, 37385, and 147493 have been determined experimentally; no information about insert sizes is available for clones 878278, 416019, and 347742. Shaded areas represent sequence similarity. Clone 27374 seems to be derived from an unspliced genomic sequence which is supported by two findings, namely lack of sequence similarity to the clones 416019 and 37385 and the presence of a CAG splice donor site at position 184 of the 5' end sequence of 27374. The orientation of the transcript cannot be deduced because directions are not consistent for overlapping clones.

levels in muscle and neuronal tissues. The expression in the kidney is restricted to the condensing mesenchyme and their derivatives, the primitive glomerular precursors that will later form nephrons (Fig. 5J–L).

No expression of the *TR2* mouse homolog could be detected in the mouse embryo, either by Northern blot analysis, or by the *in situ* experiments. The strong expression in liver and the weaker expression in kidney seems to occur in adult tissues only.

A detailed expression analysis of the *3K/fjx1* gene has already been published (Ashery-Padan et al. 1999; A. Vortkamp in prep.).

In situ hybridization with the remaining murine cDNA clones as probes (388893, 962716, 670550, 907677) did not reveal any specific expression patterns.

DISCUSSION

Contig Construction

By end probe walking we have extended a PAC contig that now consists of 201 PAC clones and covers 7.5 Mb of 11p13–14.1. Determination of the marker content and insert sizes of overlapping PACs from a highly redundant set formed the basis for exact positioning of the clones. During contig construction we found a remarkable nonrandom distribution of the PAC clones. Whereas some markers detect up to ten PACs (data not shown) we could achieve only single clone coverage in five places of the newly established clone assemblies. However, single coverage never exceeded 50 kb and was smaller than 20 kb in two cases. A fragment of genomic DNA centromeric of the *FSHB* locus (PACs 815A16/1134O24) seems to be especially difficult to clone, not only in the PAC cloning system. We could not find any additional informative clone by screening three chromosome 11 specific cosmid libraries, LLNL

(Evans and Lewis 1989), EJNAC (Heding et al. 1992), and ICRF (Nizetic et al. 1994) providing an approximate 12-fold genomic coverage with PAC ends of clones neighboring this critical area.

The newly created data fit the already existing physical maps of this region very well, reflecting the high accuracy achieved. However, an exact definition of the extent of clone overlaps is precluded at the outer edges of our region due to the paucity of *NotI* restriction sites and other mapped markers. Importantly, data obtained by fingerprinting all PAC clones strongly corroborate the hybridization results and did not provide any evidence for interstitial deletions. However, the statistical significance of results obtained with this method is necessarily limited to regions with large clone overlaps.

Gene Map of *Ilp13-14.1*

Bins defined by overlapping PAC clones are an excellent tool to accurately place ESTs and genes along the physical map of 11p13. We were able to position 15 already known genes within our PAC contig, as well as five genes that had not yet been precisely mapped (*TRAF6*, *3K/fjx1*, *PDX1*, *GA17*, and *CSTF3*) and 22 anonymous ESTs. According to the established gene map, there is a striking difference in concentration of genes in certain regions on 11p13. The 4 Mb region flanked by *RAG1/2* and *GA17* harbors 13 genes and 20 ESTs with two particular gene dense regions close to the *CSTF3* and *CD59/G2* loci. In contrast, a region of similar size telomeric of *GA17* is of low gene density and contains only seven genes (*GA17*, *WT1*, *RCN1*, *PAX6*, *239FB*, *FSHB*, and *KCNA4*) and two ESTs (Fig. 1,2). However, the established gene map may still be incomplete and in particular, genes with restricted spatial or temporal expression may be missing. Thus, gene density may change in the course of further investigations. Nevertheless, the current distribution fits the paradigm of higher gene density in R-band versus G-band regions.

RH maps are currently the most detailed source available for whole genome gene maps. It is estimated that about half of the human genes are already annotated (Deloukas et al. 1998), most of them as ESTs derived from anonymous cDNA clones. Although most of the known 11p13 genes are already included in GeneMap'98, we located three genes, *FJX1*, *GA17*, and *239FB* to our PAC contig that are not accounted for by the latest release. *FJX1* and *239FB* had been detected by

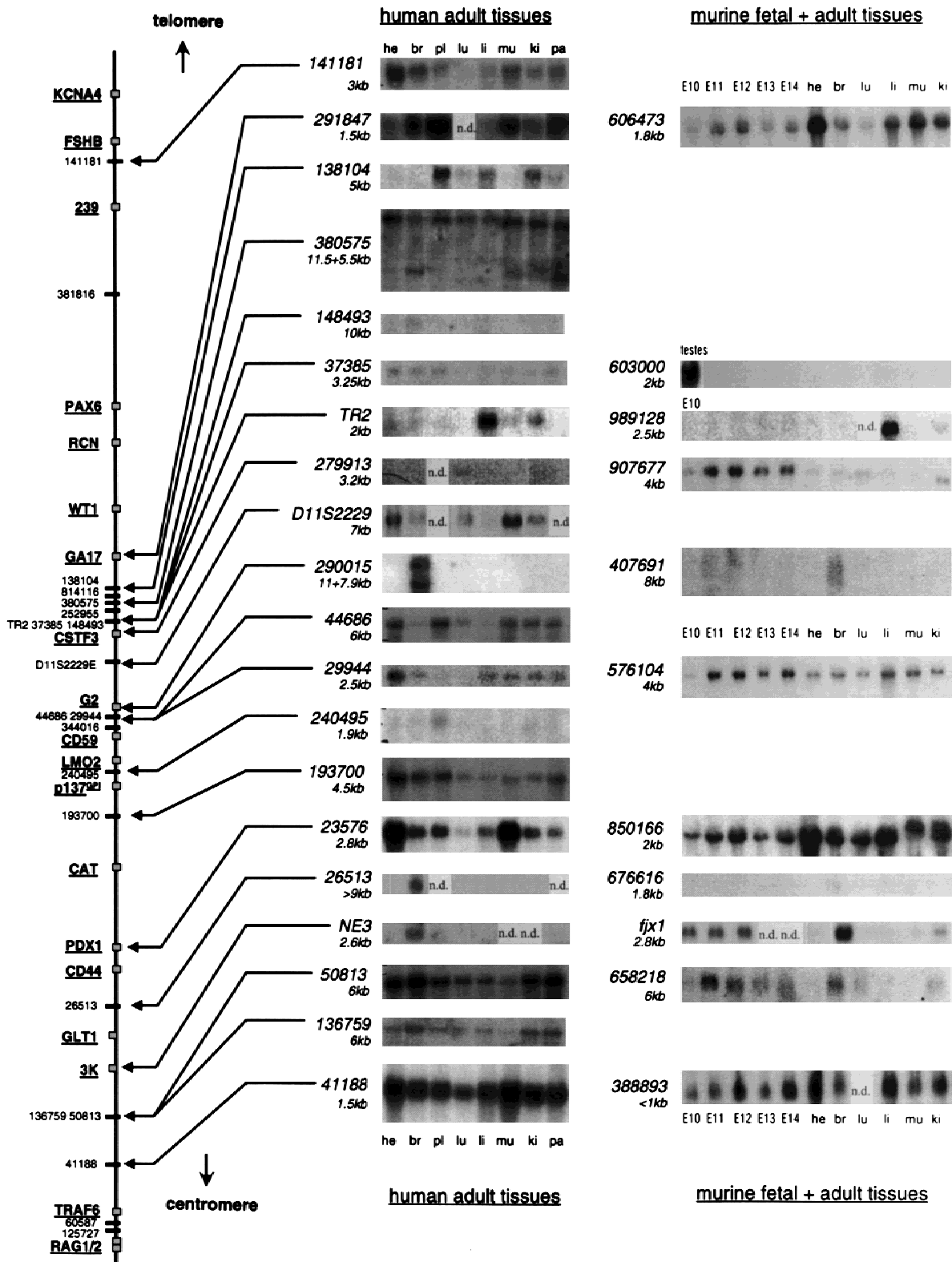


Figure 4 Northern blot analyses of the new transcripts. The results obtained by hybridization of cDNA clones to Northern blots with RNA from human (middle) or murine tissues (right) are displayed according to the genomic order of the probes. The names of the cDNA probes employed as well as the transcript sizes are indicated next to the autoradiograms. Quantitative comparison between human and murine expression is limited as the human Northern blots carry poly(A)⁺ RNAs with amounts normalized to β-actin expression, whereas an invariable amount of 20 μg of total RNA of each tissue had been applied to the murine Northern blots. RNA samples used are either derived from whole embryos of embryonic stages E10 through E14 or from adult tissues (heart, he; brain, br; placenta, pl; lung, lu; liver, li; skeletal muscle, mu; kidney, ki; pancreas, pa). For analysis of clone 603000, a testis-derived RNA sample has been added.

Table 3. Mouse Gene Probes Used for Expression Analysis

Human IMAGEclone/ gene	GenBank accession no.	Mouse I.M.A.G.E. clones	E value of blast match	Northern hybridisation	In situ hybridisation
<i>GAI7</i>	AF064603	606473	0.0	+	+
<i>pBs13</i> (37385)	X521128	603000	0.0	+	n.d.
<i>TR2</i>	TR2	989128	1e-116	+	—
<i>CSTF3</i>	U15782	907677	1e-145	+	n.s.
<i>G2</i>	U10991	407691	9e-91	+	(+)
<i>G2</i>	U10991	404066	8e-54	(+)	+
29944 5'	R14749	576104	9e-23	+	n.s.
29944 3'	R42519	670550	6e-14	—	n.s.
<i>PDX1</i>	U82328	850166	1e-119	+	+
26513	R20645	676616	3e-28	+	n.s.
<i>3K/FjX1</i>	D1153892	<i>fjx1</i>	0.0	+	n.d.
50813	H17656	658218	3e-93	+	+
136759	R34826	962716	2e-14	—	n.s.
41188	R56527	388893	2e-57	+	n.s.

Summary of mouse IMAGE cDNA clones. For the human IMAGE clone 37385 no related mouse EST clone could be identified; therefore the cDNA of the mouse protein pBs13 was used for the BLAST search. + and — point to the presence or absence of hybridization signals; brackets denote very weak signals. Hybridizations labeled as n.d. have not been performed, and n.s. alludes to nonspecific signals.

cDNA selection (Thäte et al. 1995; Schwartz et al. 1994) and *GAI7* by comparative mapping with pufferfish sequences (C. Miles and G. Elgar pers. comm.). Thus, alternative gene identification approaches are still necessary to discover the entire gene content of a region of interest.

Of 34 EST clones derived from the GeneMap'98 interval D11S4102–D11S1324 that is completely covered by our PAC contig, only 24 (71%) map to the contig clones. Especially on the level of single RH intervals, some major discrepancies occur regarding the order and distances of the markers entered (Figure 2). These inaccuracies originate from the use of the GB4 RH panel providing an average resolution of 1 Mb that is too large for fine mapping. Nevertheless, an overall concordance can be observed between the RH mapping and physical mapping approaches. Therefore, the current RH maps represent an excellent source to identify candidate genes in a distinct genomic region provided that the distortion of RH coordinates and the still preliminary gene content are taken into account. Unfortunately, the presentation of GeneMap'98 with small increments in cR values between most markers, suggests the presence of a much better resolution that cannot be gained with the panel used. The less widely used G3 panel, generated with a higher radiation dose, will certainly improve resolution.

Expression Analyses

The decision whether neighboring EST entries are part of the same gene can not always be made by mapping studies or sequence similarity searches, as cDNAs can span large genomic regions. In this respect, the com-

parison of their expression patterns provides valuable additional information.

Knowledge about transcript size and tissue specific expression of the newly mapped transcripts and their mouse homologs could be achieved rapidly by probing Northern blots with the corresponding cDNA clones. In some cases, RNA in situ hybridizations to sections of mouse embryos extend these expression profiles. The *G2* transcript and the transcript detected by the anonymous cDNA clone 658218, are both expressed specifically and almost exclusively in neuronal tissues, pointing to their potential relevance for the development and maintenance of the nervous system. The fact that *PDX1* is prominently expressed in muscle reflects the role of the PDH complex during glycolysis, namely the conversion of pyruvate to acetyl-CoA that is entered into the TCA cycle providing energy for muscle contractions. In the case of *GAI7*, the widespread expression does not yield any clues as to its function in the organism. Similarly, the analysis of the *GAI7* protein for sequence motifs only revealed the presence of a PINT domain, which itself has no ascribed function yet.

The search for homologous mouse cDNA clones was often impeded by the nature of IMAGE cDNA clones that are usually derived from oligo-dT primed libraries. This creates a strong bias towards clones representing the 3' ends of the transcripts which are often untranslated and much less conserved between species. For human and mouse cDNA probes that showed a rather weak similarity (BLAST E-values $\geq 1e^{-30}$), Northern blot analyses revealed considerably differing transcript sizes in some instances (37385/603000 and 26513/676616). This strongly suggests that the mouse

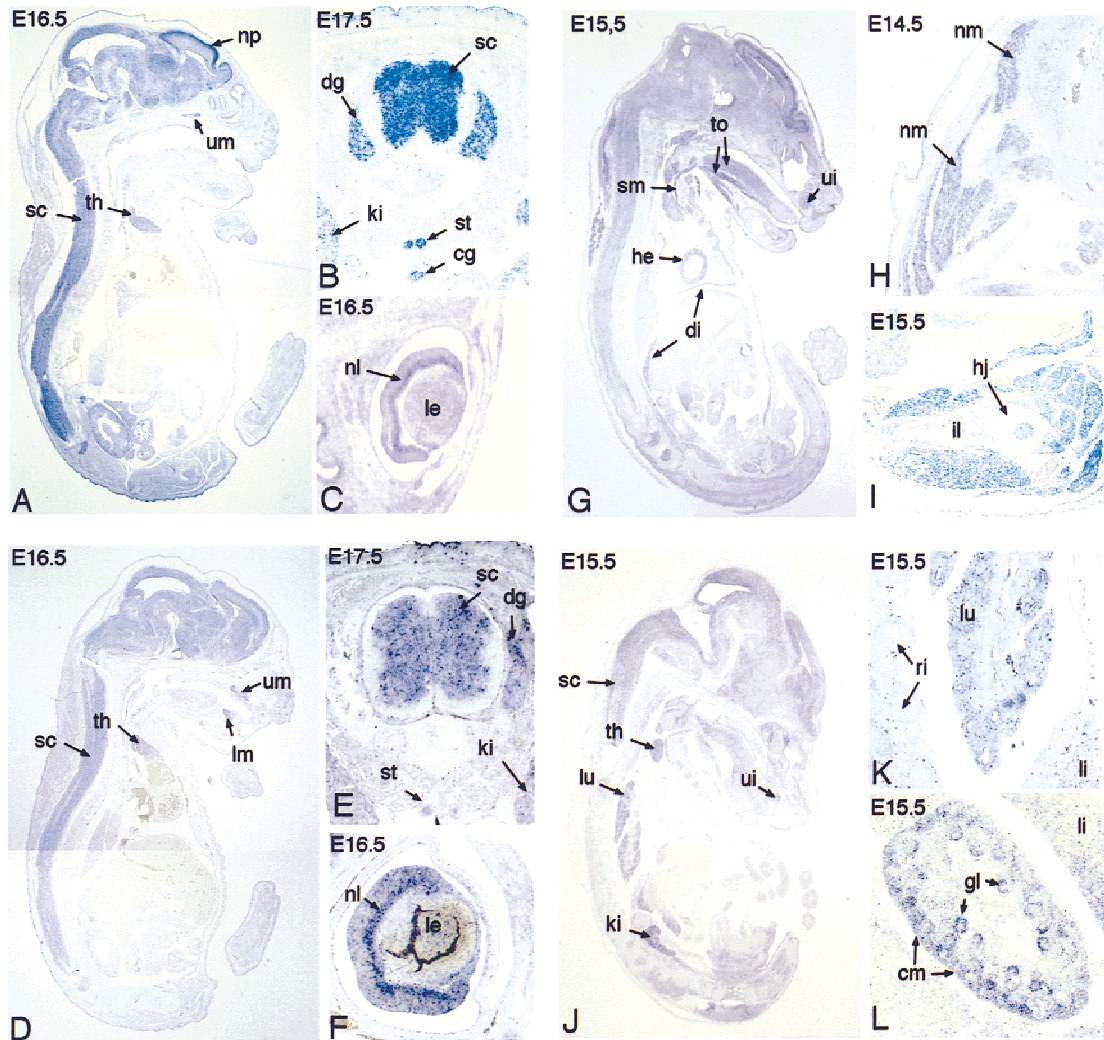


Figure 5 RNA in situ hybridizations to paraffin sections of mouse embryos, stages E14.5–17.5. (A–C) The global expression of *G2* in the nervous system is demonstrated. An even staining of spinal cord, dorsal root ganglia and brain with emphasis to the neopallial cortex can be observed as well as expression in the sympathetic trunk and the coeliac ganglion. Additionally, *G2* transcripts are seen in the inner neuroblastic area of the retina, thymus, and the primordia of the teeth. (D–F) The anonymous cDNA clone 658218 detects an expression pattern in the nervous system that is very similar to that of *G2*. (G–I) The mRNA of the X component of the PDH complex is expressed mainly in skeletal muscle, diaphragm, and the heart, but also at lower levels in brain, spinal cord, and teeth primordia. (J–L) The *GA17* mRNA is present in various tissues, mainly kidney (primitive glomeruli), lung, thymus, skeletal muscle, and teeth precursors, as well as in brain and spinal cord. Abbreviations: cg, coeliac ganglion; cm, condensing mesenchyme; dg, dorsal root ganglion; di, diaphragm; he, heart; hj, hip joint; il, iliac bone; ki, kidney; le, lens; li, liver; lm, primordium of lower molar teeth; lu, lung; nm, neck muscle; np, neopallial cortex; nr, neuroblastic layer of retina; gl, primitive glomeruli; ri, ribs; sc, spinal cord; sm, sternohyaloid muscle; st, sympathetic trunk; th, thymus gland; to, ex- and intrinsic muscles of the tongue; ui, primordium of upper incisor tooth; um, primordium of upper molar teeth.

cDNA clone identified is related, but not orthologous. Where the annotated coding sequences of genes were available for the similarity search, the identified probes generally resulted in hybridization signals consistent among all three approaches (human, as well as mouse Northern blot and in situ analyses), thus generating highly informative expression patterns.

Genomic sequencing of the present PAC contig is already under way at the Sanger Centre UK (ftp.sanger.ac.uk/pub/human/sequences/Chr_11). The knowledge of the genomic sequence will enable us in part to cir-

cumvent the lack of cDNA sequence information and to put the anonymous 3'-derived ESTs into their genomic context. The identification of murine counterparts, as well as the subsequent detailed expression analyses in the mouse, will thus become more reliable.

METHODS

Clone Libraries

PAC clones used for contig construction are derived from the RPCI-1, -3, -4, and -5 PAC libraries (Ioannou et al. 1994) and cDNA clones from the IMAGE consortium clone libraries

(Lennon et al. 1996). PAC and IMAGE clones, as well as high density PAC filters, were obtained from the Resource Centre of the German Human Genome Project, Berlin; or from a local copy of the RPCI-1 library.

Gene Specific Probes

Probes are described in Gawin et al (1995) or taken from Genome Data Base.

GenBank accession numbers are as follows: *TR2*, AJ245600; *3K*, AJ245599.

PAC Contig Construction

The methods applied for PAC contig construction have been described in detail in Niederführ et al. (1998). PAC DNAs were prepared by standard alkaline lysis followed by phenol/chloroform extraction. Clone ends were isolated according to a linker-mediated PCR protocol (Kere et al. 1992) with minor modifications. PCR products were isolated from low-melting agarose gels and directly employed for random priming without further purification. Hybridizations to colony filters of the sublibrary were carried out according to Church and Gilbert (1984) with excess human placental DNA to block repetitive sequences. Hybridization results were stored in a MS Access database. Ordering and graphical representation of the data obtained were carried out using SAM (System for Assembling Markers, Soderlund and Dunham 1995).

DNA Fingerprinting and Computational Analysis

Fingerprinting was performed according to Coulson et al. (1986) with modifications as described in Niederführ et al. (1998). Briefly, PAC DNA was cut with *HindIII*, radio labeled, and digested with *Sau3AI*. Sizes of the gel separated DNA fragments were imported into FPC allowing the identification of overlapping fingerprints. These data were transferred to 11DB, a chromosome 11 database that is based upon AceDB (<http://chr11.bc.ic.ac.uk>).

Northern Blot Analyses

Multiple human tissue Northern blots with poly(A)⁺ mRNA were obtained from Clontech or Invitrogen. Murine Northern blots contained 20 µg of total RNA prepared according to Chomczynski and Sacchi (1987) from various murine tissues or total embryos. Blots were probed with radio labeled inserts isolated from IMAGE clones or other cDNA clones as described.

In Situ Hybridization to Paraffin Sections of Mouse Embryos

In situ hybridization to paraffin sections of mouse embryos was performed as described previously (Leimeister et al. 1998). Digoxigenin-labeled riboprobes of IMAGE clones were prepared by linearization of the respective cDNA clones and subsequent transcription with T3, T7, or SP6 RNA polymerases in the presence of digoxigenin-UTP. Five-micron sections of mouse embryos embedded in paraffin were mounted on polylysine coated slides. The sections were subsequently dewaxed, rehydrated, refixed in PFA, digested with proteinase K, and fixed again. After hybridization of the riboprobes to the slides over night at 72°C, the slides were washed, RNase treated, and incubated with anti-digoxigenin antibodies at 4°C for 16–20 hr. Detection was carried out using BM-purple substrate (Boehringer).

ACKNOWLEDGMENTS

We gratefully acknowledge P. de Jong, B. Weber, and H. Stoehr for providing and transferring the RPCI-1 library, the German Resource Centre RZPD for providing PAC high density grid filters as well as PAC and IMAGE clones, C. Leimeister for helpful discussion about RNA in situ hybridization, the UK Medical Research Council and the Wellcome Trust for financial support (P.F.R.L), the Deutsche Forschungsgemeinschaft for financial support (M.G., Ge539/6), and the European Union for supporting the collaboration (Biomed 2).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Adams, M.D., J.M. Kelley, J.D. Gocayne, M. Dubnick, M.H. Polymeropoulos, H. Xiao, C.R. Merrill, A. Wu, B. Olde, R.F. Moreno et al. 1991. Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* **252**: 1651–1656.
- Aral, B., C. Benelli, G. Ait-Ghezala, M. Amessou, F. Fouque, C. Maunoury, N. Creau, P. Kamoun, and C. Marsac. 1997. Mutations in PDX1, the human lipoyl-containing component X of the pyruvate dehydrogenase-complex gene on chromosome 11p1, in congenital lactic acidosis. *Am. J. Hum. Genet.* **61**: 1318–1326.
- Ashery-Padan, R., G. Alvarez-Bolado, B. Klamt, M. Gessler, and P. Gruss. 1999. *Fjx1*, the murine homologue of the drosophila four-jointed gene, codes for a putative secreted protein expressed in restricted domains of the developing and adult brain. *Mech. Dev.* **80**: 213–217.
- Beppler, G. and M.A. Garcia-Blanco. 1994. Three tumor-suppressor regions on chromosome 11p identified by high-resolution deletion mapping in human non-small-cell lung cancer. *Proc. Natl. Acad. Sci.* **91**: 5513–5517.
- Call, K.M., T. Glaser, C.Y. Ito, A.J. Buckler, J. Pelletier, D.A. Haber, E.A. Rose, A. Kral, H. Yeger, W.H. Lewis et al. 1990. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell* **60**: 509–520.
- C. elegans Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282**: 2012–2018.
- Church, G.M., and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci.* **81**: 1991–1995.
- Chomczynski, P. and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**: 156–159.
- Coulson, A., J. Sulston, S. Brenner, and J. Karn. 1986. Towards a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **83**: 7821–7825.
- Deloukas, P., G.D. Schuler, G. Gyapay, E.M. Beasley, C. Soderlund, P. Rodriguez-Tome, L. Hui, T.C. Matise, K.B. McKusick, J.S. Beckmann et al. 1998. A physical map of 30,000 human genes. *Science* **282**: 744–746.
- Evans, G.A. and K.A. Lewis. 1989. Physical mapping of complex genomes by cosmid multiplex analysis. *Proc. Natl. Acad. Sci.* **86**: 5030–5034.
- Fantes, J., B. Redeker, M. Breen, S. Boyle, J. Brown, J. Fletcher, S. Jones, W. Bickmore, Y. Fukushima, M. Mannens and et al. 1995. Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum. Mol. Genet.* **4**: 415–422.
- Gawin, B., B. Klamt, A. Konig, C. Thäte, D. Le Paslier, I. Chumakov, R. Bhogal, G. Zehetner, G. Bruns, and M. Gessler. 1995. An integrated YAC clone contig for the WAGR region on human chromosome 11p13–p14.1. *Genomics* **30**: 37–45.
- Gessler, M., A. Poustka, W. Cavenee, R.L. Neve, S.H. Orkin, and G.A.

- Bruns. 1990. Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. *Nature* **343**: 774–778.
- Glaser, T., C.A. Jones, W.H. Lewis, K. Call, E.A. Rose, A. Buckler, C. Ito, and D.E. Housman. 1989. The ultrafine structure of the WAGR complex. *Cytogenet. Cell. Genet.* **51**: 1005.
- Gyapay, G., K. Schmitt, C. Fizames, H. Jones, N. Vega-Czarny, D. Spillet, D. Muselet, J. F. Prud'Homme, C. Dib, C. Auffray, J. Morissette, J. Weissenbach, and P. N. Goodfellow. 1996. A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5**: 339–346.
- Heding, I.J., A.C. Ivens, J. Wilson, M. Strivens, S. Gregory, J.M. Hoovers, M. Mannens, B. Redeker, D. Porteous, V. van Heyningen et al. 1992. The generation of ordered sets of cosmid DNA clones from human chromosome region 11p. *Genomics* **13**: 89–94.
- Hillier, L.D., G. Lennon, M. Becker, M.F. Bonaldo, B. Chiapelli, S. Chissoe, N. Dietrich, T. DuBuque, A. Favello, W. Gish et al. 1996. Generation and analysis of 280,000 human expressed sequence tags. *Genome Res* **6**: 807–828.
- Ioannou, P.A., C.T. Amemiya, J. Garnes, P.M. Kroisel, H. Shizuya, C. Chen, M.A. Batzer, and P.J. de Jong. 1994. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* **6**: 84–89.
- Kere, J., R. Nagaraja, S. Mumm, A. Ciccodicola, M. D'Urso, and D. Schlessinger. 1992. Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts. *Genomics* **14**: 241–248.
- Kiechle-Schwarz, M., T. Bauknecht, U. Karck, F. Kommos, A. du Bois, and A. Pfeleiderer. 1994. Recurrent cytogenetic aberrations and loss of constitutional heterozygosity in ovarian carcinomas. *Gynecol. Oncol.* **55**: 198–205.
- Leimeister, C., A. Bach, and M. Gessler. 1998. Developmental expression patterns of mouse sFRP genes encoding members of the secreted frizzled related protein family. *Mech. Dev.* **75**: 29–42.
- Lennon, G., C. Auffray, M. Polymeropoulos, and M.B. Soares. 1996. The IMAGE Consortium: An integrated molecular analysis of genomes and their expression. *Genomics* **33**: 151–152.
- Ling, M., G. McEachern, A. Seyda, N. MacKay, S.W. Scherer, S. Bratinova, B. Beatty, M.L. Giovannucci-Uzielli, and B.H. Robinson. 1998. Detection of a homozygous four base pair deletion in the protein X gene in a case of pyruvate dehydrogenase complex deficiency. *Hum. Mol. Genet.* **7**: 501–505.
- Miles, C., G. Elgar, E. Coles, D.J. Kleinjan, V. van Heyningen, and N. Hastie. 1998. Complete sequencing of the Fugu WAGR region from WT1 to PAX6: Dramatic compaction and conservation of synteny with human chromosome 11p13. *Proc. Natl. Acad. Sci.* **95**: 13068–13072.
- Niederführ, A., Hummerich, H., Gawin, B., Boyle, S., Little, P.F.R., and Gessler, M. 1998. A sequence-ready 3-Mb PAC contig covering 16 breakpoints of the Wilms Tumor/Aniridia region of human chromosomes 11p13. *Genomics* **53**: 155–163.
- Nizetic, D., S. Monard, B. Young, F. Cotter, G. Zehetner, and H. Lehrach. 1994. Construction of cosmid libraries from flow-sorted human chromosomes 1, 6, 7, 11, 13, and 18 for reference library resources. *Mamm. Genome* **5**: 801–802.
- Pennisi, E. 1999a. Academic sequencers challenge Celera in a sprint to the finish. *Science* **283**: 1822–1823.
- Pennisi, E. 1999b. Fruit fly researchers sign pact with Celera. *Science* **283**: 767.
- Schedl, A., A. Ross, M. Lee, D. Engelkamp, P. Rashbass, V. van Heyningen, and N.D. Hastie. 1996. Influence of PAX6 gene dosage on development: Overexpression causes severe eye abnormalities. *Cell* **86**: 71–82.
- Scholz, H., S.A. Bossone, H.T. Cohen, U. Akella, W.M. Strauss, and V.P. Sukhatme. 1997. A far upstream cis-element is required for Wilms' tumor-1 (WT1) gene expression in renal cell culture. *J. Biol. Chem.* **272**: 32836–32846.
- Shipman, R., P. Schraml, M. Colombi, and C.U. Ludwig. 1998. Allelic deletion at chromosome 11p13 defines a tumour suppressor region between the catalase gene and D11S935 in human non-small cell lung carcinoma. *Int. J. Oncol.* **12**: 107–111.
- Shipman, R., P. Schraml, M. Colombi, G. Raefle, and C.U. Ludwig. 1993. Loss of heterozygosity on chromosome 11p13 in primary bladder carcinoma. *Hum. Genet.* **91**: 455–458.
- Soderlund, C., and I. Dunham. 1995. SAM: A system for iteratively building marker maps. *Comput. Appl. Biosci.* **11**: 645–655.
- Thäte, C., J. Pongratz, A. König, B. Klamt, S. Tsaoussidou, M. Higgins, T. Shows, C. Jones, and M. Gessler. 1995. CpG island clones for chromosome 11p: A resource for mapping and gene identification. *Mamm. Genome* **6**: 421–425.
- Ton, C.C., H. Hirvonen, H. Miwa, M.M. Weil, P. Monaghan, T. Jordan, V. van Heyningen, N.D. Hastie, H. Meijers-Heijboer, M. Drechsler et al. 1991. Positional cloning and characterization of a paired box- and homeobox- containing gene from the aniridia region. *Cell* **67**: 1059–1074.
- Turleau, C., J. de Grouchy, M.F. Tournade, M.F. Gagnadoux, and C. Junien. 1984. Del 11p/aniridia complex: Report of three patients and review of 37 observations from the literature. *Clin. Genet.* **26**: 356–362.
- Vandamme, B., W. Lissens, K. Amfo, P. De Sutter, C. Bourgain, E. Vamos, and J. De Greve. 1992. Deletion of chromosome 11p13–11p15.5 sequences in invasive human ovarian cancer is a subclonal progression factor. *Cancer Res.* **52**: 6646–6652.
- Wadman, M. 1999. Human Genome Project aims to finish "working draft" next year. *Nature* **398**: 177.

Received July 8, 1999; accepted in revised form August 25, 1999.