

Cloning and Characterization of 13 Novel Transcripts and the Human RGS8 Gene from the 1q25 Region Encompassing the Hereditary Prostate Cancer (HPC1) Locus

Raman Sood,^{*,1,2} Tom I. Bonner,^{†,1} Izabela Makalowska,[‡] Dietrich A. Stephan,^{*} Christiane M. Robbins,^{*} Tim D. Connors,[§] Sharon D. Morgenbesser,[§] Kui Su,[§] Mezbah U. Faruque,^{*} Heather Pinkett,^{*} Chris Graham,^{*} Andreas D. Baxeavanis,[‡] Kathy W. Klinger,[§] Greg M. Landes,[§] Jeffrey M. Trent,^{*} and John D. Carpten^{*}

^{*}Cancer Genetics Branch and [‡]Genome Technology Branch, National Human Genome Research Institute, and [†]Laboratory of Genetics, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892; and [§]Genzyme Genetics Corporation, Framingham, Massachusetts 01701

Received November 6, 2000; accepted January 12, 2001

The aim of this study was to develop a saturated transcript map of the region encompassing the HPC1 locus to identify the susceptibility genes involved in hereditary prostate cancer (OMIM 176807) and hyperparathyroidism–jaw tumor syndrome (OMIM 145001). We previously reported the generation of a 6-Mb BAC/PAC contig of the candidate region and employed various strategies, such as database searching, exon-trapping, direct cDNA hybridization, and sample sequencing of BACs, to identify all potential transcripts. These efforts led to the identification and precise localization on the BAC contig of 59 transcripts representing 22 known genes and 37 potential transcripts represented by ESTs and exon traps. Here we report the detailed characterization of these ESTs into full-length transcript sequences, their expression pattern in various tissues, their genomic organization, and their homology to known genes. We have also identified an *Alu* insertion polymorphism in the intron of one of the transcripts. Overall, data on 13 novel transcripts and the human RGS8 gene (homologue of the rat RGS8 gene) are presented in this paper. Ten of the 13 novel transcripts are expressed in prostate tissue and represent positional candidates for HPC1. © 2001

Academic Press

INTRODUCTION

Prostate cancer (OMIM 176807) is a complex disease involving multiple susceptibility loci as evident from genome-wide linkage scans performed over the past 4 years following the identification of the HPC1 locus by our laboratory (Smith *et al.*, 1996). Several loci, including three on chromosome 1 designated HPC1 (1q24–q31, Smith *et al.*, 1996; Cooney *et al.*, 1997; Neuhausen *et al.*, 1999), PcaP (1q42, Berthon *et al.*, 1998), and CAPB (1p36, Gibbs *et al.*, 1999), the HPCX locus on the X chromosome (Xu *et al.*, 1998), and one each on chromosomes 2q, 12p, 15q, 16p, and 16q (Suarez *et al.*, 2000), and 20q13 (Berry *et al.*, 2000), have been implicated to contain potential prostate cancer susceptibility genes. Recently Rebbeck *et al.* (2000) reported the association between two common missense variants in the HPC2/ELAC2 gene and prostate cancer risk. Despite extensive molecular studies, no gene has been identified yet that would help elucidate the underlying mechanisms responsible for hereditary prostate cancer. It is difficult to draw any conclusions about the nature of the genetic defect involved in HPC1-associated prostate cancer cases based on existing cytogenetic data. There is no evidence of gross cytogenetic rearrangements involving the HPC1 candidate region. Furthermore, loss of heterozygosity studies at the HPC1 locus have shown that the loss of wildtype allele, which is a hallmark of classic tumor suppressor genes, is very rare in both familial and sporadic cases of prostate cancer (Latil *et al.*, 1997; Dunsmuir *et al.*, 1998; Ahman *et al.*, 2000). There are no reports of loss at 1q24–q31 by comparative genomic hybridization (CGH) in sporadic prostate cancer cases (Visakorpi *et al.*, 1995; Sattler *et al.*, 1999). However, there has been one report of gain of 1q by CGH, suggesting the possible involvement of an oncogene (Cher *et al.*, 1996).

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AF288389–AF288399, AF297014–AF297023, AF338436, and AF312863–AF312865.

¹ These authors contributed equally to this work.

² To whom correspondence should be addressed at NHGRI, NIH, Building 36, Room 3D05, 9000 Rockville Pike, Bethesda, MD, 20892. Telephone: (301) 435-5746. Fax: (301) 435-5465. E-mail: rsood@nhgri.nih.gov.



Our laboratory has focused on the identification of the HPC1 gene using a positional cloning strategy. We have developed a sequence-ready BAC/PAC contig covering 6 Mb of the HPC1 candidate region, which served as a template to identify and map 22 known genes and 37 potential transcripts (Carpten *et al.*, 2000). Several known genes and ESTs were identified from the human gene map and mapped to the contig at high resolution. In addition, we used BAC and PAC clones as templates to identify unmapped ESTs and potential coding sequences by exon-trapping, shotgun sequencing, and direct screening of cDNA libraries. All ESTs and exon traps representing potential transcripts were characterized for their expression profiles using Northern blot analysis. We have extended several partial cDNA sequences into near full-length clones through a combination of gene-trapper (LTI) and RACE (Clontech) techniques, in addition to assembly and editing of EST sequences from GenBank. The characterization of these potential transcripts will provide not only candidate genes for HPC1 but also HRPT2 (hyperparathyroidism-jaw tumor syndrome; OMIM 145001), which has been mapped to the same chromosomal region (Teh *et al.*, 1996; Hobbs *et al.*, 1999; Williamson *et al.*, 1999).

MATERIALS AND METHODS

Transcript identification and mapping. Details of the clone-based physical map and identification of transcripts are described elsewhere (Carpten *et al.*, 2000). Briefly, transcripts were identified by database searches utilizing human gene maps (<http://www.ncbi.nlm.nih.gov/genemap/>), exon-trapping of BAC clones representing the minimal tiling path, sample sequencing of selected BAC clones followed by computational analysis, and also direct hybridization of high-density nylon filters arrayed with IMAGE cDNA clones (Research Genetics) with radiolabeled BAC inserts. For details of these methods, please refer to our earlier publication (Carpten *et al.*, 2000). All potential transcripts were precisely placed on the physical map either by PCR (polymerase chain reaction) or by Southern blotting of BAC clones followed by hybridization with inserts from available cDNA clones.

Northern analysis. Northern blots with mRNA from multiple human tissues including prostate (MTN4) were purchased from Clontech. To determine the transcript size(s) and tissue distribution pattern of potential transcripts, MTN4 blots were hybridized with either gel-purified inserts from available IMAGE cDNA clones or PCR products generated from cDNA libraries (Clontech). The latter was used if no IMAGE clone was available or if available clones contained a repeat element such as *Alu* or L1. The probes were labeled with [³²P]dCTP by random priming (Stratagene) following the manufacturer's directions. Hybridization was carried out at 42°C overnight in Hybrisol 1 hybridization buffer (Intergen) followed by stringent washing in 0.1%SDS, 0.1× SSC at 65°C. Autoradiography was performed for the appropriate time period at -80°C with a BioMax intensifying screen and film.

cDNA library screening by Gene Trapper. Longer cDNA clones for potential transcripts were identified using the Gene Trapper cDNA Positive Selection System (Life Technologies) following the manufacturer's directions. Briefly, oligonucleotides designed from partial cDNA sequences representing exon traps or ESTs were synthesized and PAGE-purified (Lofstrand Labs). These oligonucleotides were then biotinylated and hybridized to single-stranded DNA from adult and fetal brain, liver, spleen, leukocytes, kidney, heart, lung, and testis cDNA libraries. Hybrids were captured using paramagnetic streptavidin beads, and double-stranded cDNA was reconstructed

using another gene-specific oligonucleotide downstream of that used in the initial hybridization reaction. Positive clones were identified by transformation of resulting double-stranded DNA into *Escherichia coli*, followed by PCR amplification of recombinant clones using gene-specific primers.

Rapid amplification of cDNA ends (RACE). 5'-RACE was performed using the Marathon-Ready RACE system (Clontech) according to their specifications. Marathon-Ready cDNA libraries derived from fetal and adult brain, prostate, testis, liver, leukocytes, and spleen were used as templates for RACE reactions. Typically, three antisense gene-specific primers were designed from the partial cDNA sequence. Adaptor primer AP1 was used in the primary RACE reaction with each of the two most 3' primers. RACE reactions were purified from unincorporated nucleotides and primers using the QIAquick PCR purification kit (Qiagen) and then used as template for secondary RACE reactions using the AP2 adaptor primer and nested gene-specific primers. Products were cloned using the TOPO TA cloning kit (Invitrogen).

Clone characterization. All IMAGE clones, positive Gene-Trapper clones, and RACE clones were processed as follows: DNA was extracted from a 3- to 5-ml overnight bacterial culture using the Wizard Plus Minipreps DNA Purification System (Promega). Clones were sequenced using primers from both ends of the vector followed by walking primers designed from the initial sequence data. DNA sequencing reactions were performed using ~0.5 µg DNA per reaction and the PE Biosystems FS+ and Big Dye Terminator chemistry (Robbins *et al.*, 1996) and were analyzed on a PE Biosystems 377 XL DNA Sequencer.

Computer analysis. Sequences were assembled into contigs and analyzed for open reading frames using Sequencher, v3.1 (Gene Codes). When available (<ftp://genome.wustl.edu/pub/gsc/est/>), trace data for ESTs were incorporated and manually edited to remove base-calling errors. Sequence homology to previously characterized DNA and protein sequences in the nr and dbEST sections of GenBank was determined using the BLAST algorithm (Altschul *et al.*, 1990) accessible from the NCBI home page (<http://www.ncbi.nlm.nih.gov/>). A reiterative process of assembly and BLAST searches allowed extension of many ESTs well beyond the ends of Unigene clusters. We have submitted the sequences of all characterized transcripts to GenBank, and the accession numbers are provided for further reference (Table 2). Protein sequences predicted from the cDNAs were analyzed for motifs using Interpro (<http://www.ebi.ac.uk/interpro/>).

RESULTS

We used a combination of exon-trapping, cDNA library screening with BAC inserts, and shotgun sample sequencing in addition to the information available in the public databases, particularly the human gene maps (Deloukas *et al.*, 1998), to identify transcribed sequences from the region encompassing the HPC1 locus at 1q25. These efforts identified 22 known genes and 37 ESTs that were localized to the 6-Mb BAC/PAC contig (Carpten *et al.*, 2000). Further characterization of the ESTs was performed as follows: Unigene clusters were used to assemble maximum available sequence information, and IMAGE clones for the largest available clone in each cluster were purchased (Research Genetics) and sequenced to verify the clone and to obtain the sequence of the entire insert. The sequences were searched for a putative polyadenylation signal (AATAAA or with minor variation as described by Beaudoin *et al.*, 2000) and database homologies to any known genes or domains indicative of potential coding sequence. Since our tissue of interest has been pros-

TABLE 1
Sequence Features of ESTs Not Expressed in Tissues on MTN4 Blot

Name	Accession No.	Unigene Cluster ID	Sequenced length (bp)	Poly(A) signal	Characteristics of available sequence
AA757084	AA757084	None	486	Yes (AGTAAA)	Single EST, no other hits
e10	T97314	Hs. 18033	1300 in 3 fragments	Yes	Single EST
e7	AF075081	Hs. 36300		No	No ORF
AA431804	AA431804	None	546	5' end sequenced	Overlaps with a mouse EST
e9	T79647	None	513	No	Single EST, no other hits
e4	R48958	Hs. 221535	940 in 2 fragments	No	No ORF
e13	H92337	None		434	No
ge27	AA657406	None	439	No	Identified by an exon trap, intronic sequence on both sides of the exon
e5	M78545	None	276	5' end sequenced	Single EST, no other hits
AA744170	AA744170	None	305	No	Single EST, no other hits
IC-145842	R79368	Hs. 144508	930	Yes	5' end contains <i>Alu</i> sequence May represent alternate 3' end of PLA2
IC-123336	R00375	None	710	No	Repeat sequence
Z39494	Z39494	Hs. 65765	316	No	No other hits
M78285	M78285	None	442	Yes	Single EST, no other hits
M78354	M78354	None	352	Yes	Single EST, no other hits

tate, we hybridized inserts from these clones to MTN4 Northern blots (Clontech) to determine both the size and the tissue-specificity of the expected transcripts. ESTs were prioritized for further characterization based on a positive signal on the Northern blots and the presence of a polyadenylation signal, indicative of the ESTs representing authentic cDNAs.

Of the 37 candidate ESTs, 15 failed to meet our criteria and were not characterized further (Table 1). Apart from background signals or smear, no clear bands were observed on MTN4 Northern blots for any of these ESTs. Furthermore, their initial characterization showed most of them to be represented by a single cDNA clone. Eight of these ESTs also lacked a polyadenylation signal, thereby suggesting that they may have been primed on the A-rich stretches of intronic sequences, representing artifacts of cDNA cloning (Wolfsberg and Landsman, 1997). We were particularly interested in ge27, since it was identified by an exon trap and the EST is derived from a prostate cDNA library. However, repeated attempts at 5'-RACE yielded only products that matched genomic sequence. These data indicate that a large proportion of these ESTs represent artifacts of cDNA cloning. However, the remaining few might represent transcripts expressed in tissues other than those on MTN4 Northern blots and, therefore, need to be analyzed further, particularly for their expression in parathyroid gland and kidney, to be considered as candidates for HRPT2 syndrome.

Of the remaining 22 ESTs, data on the detailed characterization of 19 transcripts are presented below. KIAA0959/e3 was found to represent the human homologue of mouse RGL, a protein involved in the Ras signaling pathway, and has been described elsewhere (Sood *et al.*, 2000). Similarly, Z47, an 18-kb cDNA

representing a human homologue of the *Caenorhabditis elegans* hemiceptin gene him-4 has been characterized extensively and will be described elsewhere (manuscript in preparation). The EST e19 was found to encode the 3'-UTR of the voltage-dependent calcium channel α -1E subunit (GenBank Accession No. L29384), a known gene that we had previously mapped to the same set of overlapping BAC clones.

Two methods of transcript extension, the Gene-Trapper Positive cDNA Selection System (LTI) and 5'-RACE (Clontech), were used in extending these 19 ESTs to full-length transcripts. High GC 5'-RACE has also been used in some cases to identify GC-rich 5' ends. The method of choice for cDNA extension was based upon the results of screening a panel of Marathon-Ready cDNA libraries (Clontech) and superscript cDNA libraries (LTI) from a variety of tissues. Resulting clones were sequenced and assembled with the data from the Unigene cluster using Sequencher. The cDNA and deduced amino acid sequences of all transcripts were used to search for any homologous sequences in GenBank using the BLAST programs. These analyses showed that 13 of the 19 transcripts are novel, 2 are homologous to previously cloned genes, 2 are redundant with others, and 2 most likely represent artifacts. Figure 1 shows the position of these transcripts relative to one another and genetic markers from the 1q25 region. The orientation of these transcripts was determined either by PCR using primers from 5' and 3' ends of the cloned sequence against the BAC contig or by alignment of the cDNA sequences with genomic sequences and is shown by the arrow. Transcripts c190 and c78 as well as the ESTs Z39494, M78285, and M78354 map telomeric of D1S1642. Table 2 lists various characteristics of these transcripts, including the expected size from Northern blot analy-

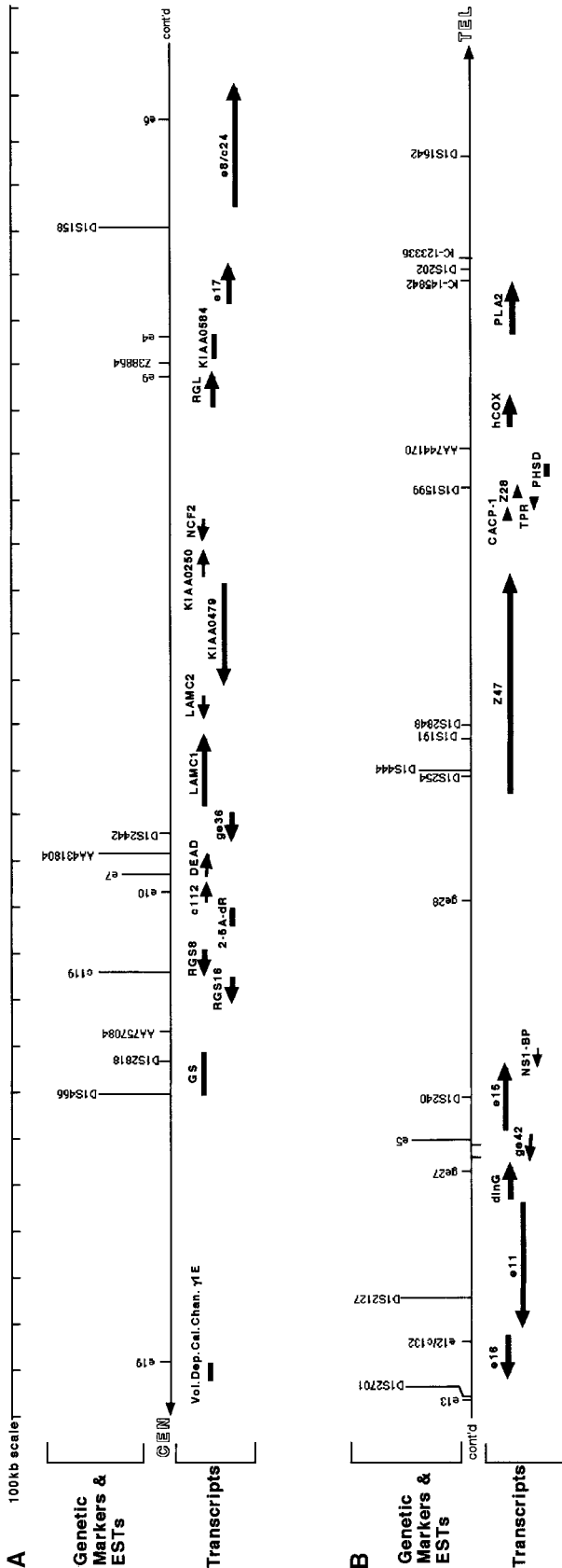


FIG. 1. Transcript map of the 1q25 region in centromeric to telomeric orientation shown in two tiers (A and B). Genetic markers and 100-kb scale are given as the reference points. Transcripts are shown as horizontal lines with or without an arrow indicating the direction of transcription. All other ESTs are shown as STS markers along with the genetic markers.

sis, cloned size and length of the open reading frame (ORF), database similarities to known genes from various organisms, and genomic organization. The transcripts are listed in centromeric to telomeric order.

Expression Profiles of Novel Transcripts

The expression profiles of the novel genes were examined with eight different human tissues, including prostate. Representative Northern blots showing the transcript size and tissue distribution pattern for eight transcripts are presented in Fig. 2. The data are also summarized in Table 2. Ten transcripts showed expression in prostate; however, none revealed a prostate-specific expression pattern. Seven of the 10 were found to be expressed to varying degrees in all the tissues examined, suggesting that these genes are expressed ubiquitously. z28 was found to be expressed in three alternative forms in all tissues except peripheral blood leukocytes, with the largest form (4.2 kb) being the most abundant. Hybridization signals for 3 transcripts were observed in testis only, with ge36 being very strong, whereas e15 and KIAA0479/e20/c37 were detected at low levels. Low-level expression in testis is often seen even for transcripts specifically expressed in other tissues. Therefore, e15 and e20 may not be testis-specific genes but expressed in tissues not examined in this study. The fact that the KIAA0479 sequence was derived from a cDNA clone originating from a brain cDNA library suggests that KIAA0479/e20/c37 may be a brain-specific transcript. Hybridization of MTN1 Northern blot with a probe derived from this transcript revealed that KIAA0479/e20/c37 has strong expression in brain as well as weak expression in placenta (data not shown).

Sequence Characteristics of Novel Transcripts

Sequence analysis has shown 13 of the 19 transcripts to be novel, with varying degrees of homology to proteins from other species. Sequencing of random cDNA clones with large inserts (RIKEN project and KIAA series initiated by Ohara *et al.*, 1997) has also identified partial sequences for several of our transcripts. These transcripts are indicated by names comprising database name/est or contig number assigned to all our ESTs as an identifier (Table 2), e.g., KIAA0584/c33/c47 (Nagase *et al.*, 1998), KIAA0250/c48 (Nagase *et al.*, 1996), and KIAA0479/e20/c37 (GenBank Accession No. AB007948). In all cases, our data provide either additional coding sequence at 5' ends or detailed expression profiles. New accession numbers were obtained only if our sequence provided additional data; e.g., the e20/c37 sequence is longer than the KIAA0479 sequence and provides additional data at the 5' end.

Our z28 cDNA (not submitted to GenBank) corresponds to two overlapping cDNA sequences, AK000512 and AK000171, which differ in their utilization of polyadenylation sites and an insertion of A near the 3' end of the coding sequence. There are eight As in the

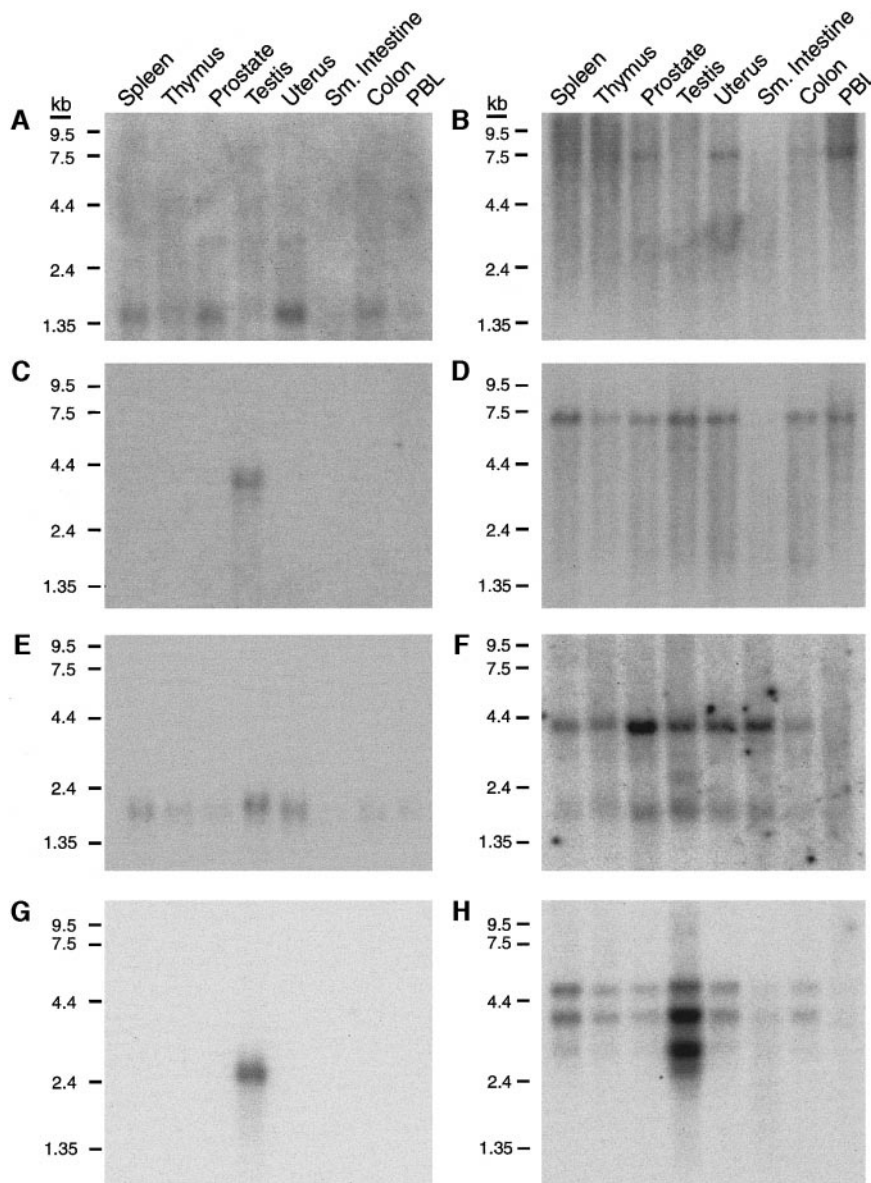


FIG. 2. Northern blot analysis of eight transcripts from the HPC1 candidate region. Multiple tissue Northern blots (Clontech) containing poly(A)⁺ mRNA of eight adult human tissues, as indicated at the top of each lane, were hybridized with ³²P-labeled cDNA probes specific for e8 (A), e11 (B), e15 (C), e16 (D), e17 (E), z28 (F), ge36 (G), and ge42 (H). Size markers are indicated on the left side of each blot. Blots were autoradiographed for 1–5 days at –80°C.

AK000171 sequence versus nine As in the AK000512 sequence. The (A)₈ allele gives an open reading frame that extends 360 bases farther 3' than that of the (A)₉ allele. We analyzed DNA from 12 individuals by sequencing an amplicon of 414 bp containing the presumed polymorphic site and failed to observe any chromosome with the (A)₉ allele. Analysis of the genomic sequences in the htgs database identifies two BAC clones (GenBank Accession Nos. AL162722 and AL162723), both with the (A)₈ allele. The only other sequence with the (A)₉ allele is an EST sequence (GenBank Accession No. AA085997). These data suggest that the (A)₉ allele represents either a rare allele or a sequencing error/cloning artifact.

We have cloned the complete coding region for e8/

c24, e20, c78, z28/c7, ge36, and ge42, as evidenced by the presence of in-frame termination codons upstream of the first ATG in these six transcripts. Furthermore, Northern blot data indicate that almost the complete coding sequence has been obtained for e11, e15, and e17. The ATG coding for the initiating methionine lies in the optimal context for initiation of translation as suggested by Kozak (1996) in these three transcripts, further supporting the conclusion that we have identified the complete coding sequence for these transcripts.

Known Genes

Homology searches with one of our ESTs (c78) identified a known gene, B3GalT2 (GenBank Accession No.

TABLE 2
Sequence Data and Expression Profiles of Transcripts

Name	HUGO nomenclature	Unigene Cluster ID	Accession No.	Cloned length (bp)	ORF (aa)	Expression data		Homology and putative function	Genomic organization
						Size (kb)	Tissue specificity		
RGS8	RGS8	None	AF297015	765	198	6.5	Brain	Rat RGS8; regulation of G protein signaling	AL138776, AL158215, & AL353778; >6 exons
c119	RGS8	Hs. 20982	AF297014	1,496	None	Not determined		3'-UTR of RGS8	AL138776, 3'-UTR of RGS8
c112	Clorf13	Hs. 23756	AF338346	2,609	320	5	All tissues (weak)	Novel; 40% identity to N-acetylneuraminidase (AAF94925)	AL355999, AC016364, AL133333,
ge36	Clorf14	None	AF288397 (alt) & AF288398	2,022 2,363	534 725	2.5	Testis (very strong)	Novel; 30% identity to mouse Shc binding protein (AAD01613)	AL135843; 13 exons AF297016-AF297023 for exons with flanking intron sequences
KIAA0479/e20/c37	Clorf15	Hs. 158244	AF288395	5,696	307	6.5	Testis (weak)	Novel; distant homology to yeast membrane protein	AC004795, AL139013; 11 exons
c48/KIAA0250	Clorf16	Hs. 15087	D87437	5,082	802	6	All tissues	Novel; no homology detected	AL357876; 17 exons
c33/c47/KIAA0584	Clorf17	Hs. 106794	AF288389	5,174	626	Not determined		Novel; 53% identity to cerebral cell adhesion molecule (AF177203)	AL157950 & AL157943; 12 exons
Z38864	Clorf18	Hs. 26835			Not determined			3' end of c33/c47/KIAA0584	
e17	Clorf19	Hs. 128769	AF288394	1,961	>197	2.1	All tissues, strong in testis & uterus	Novel; no homology detected	AL158011, AL157943; 5 exons
e6	Clorf20	A009C004	AF288396	2,517	None	4.5	Spleen, testis, uterus	No ORF, may not represent a true transcript	AL078645; in the intron of e8
e8/c24	Clorf21	Hs. 16769	AF312864	10,215	121	11, 1.5	Spleen, prostate, testis, uterus PBL	Novel; identical to partial predicted protein CAB92738	AL078645, AL358152; >6 exons
e16	Clorf22	Hs. 25264	AF288393	6,298	>906	7.5	All tissues	Novel; identical to partial predicted protein CAC03447; 48% identity to CG5682 gene product (Drosophila); 29% identity to α -mannosidase in 5' region	AL096819; 20 exons
c132/e12	Clorf23	Hs. 117788	Not submitted	7,100	None	7.5	All tissues	Lies in the introns of e16 on opposite strand	AL096819; in the intron of e16
e11	Clorf24	Hs. 48778	AF288391	6,919	>992	7.5	Prostate, testis, uterus	DNAJ-1 motif, identical to partial predicted protein CAB75864	AL109865, AL109956, AL096819, AL136086; 14 exons

ge42	Clorf25	None	AF288399	4,326	730	4,8,3,6,2,8 minor	All tissues	Novel; 20% identity to dimethylguanosine tRNA methyltransferases, identical to partial predicted proteins CAC00613 and CAC00614	AL109865; 15 exons plus 2 alternative 5' exons
el5	Clorf26	Hs. 113928	AF288392 AK000128	3,604	>900	4.1	Testis (weak)	Novel	AL136133 & AL356273 14 exons
z28/c7	Clorf27	Hs. 69388	AK000512 AK000171	1,791 2,131	335	4,2,1,9,1,7	All tissues except PBL	25% identity to yeast protein CAA99372 over 200-aa segment	AL162722 & AL162723 14 exons
c190	Clorf28	Hs. 5722	AF312865	2,318	531	Not determined		22% identity to <i>C. elegans</i> odorant receptor accessory protein	>15 exons
c78/B3GalT2	B3GALT2	Hs. 181353	AF288390	3,215	5,5,3,8		All tissues (weak)	Novel; 57% identity to <i>Drosophila</i> predicted gene AAF54331	AL390863, AL139133; 2 exons

Y15014), which belongs to a family of human β -3-galactosyltransferases involved in the synthesis of type 1 carbohydrate chains (Kolbinger *et al.*, 1998). It should be noted that the 5'-UTR sequence of this transcript identified by us (GenBank Accession No. AF288390) is different from the published 5'-UTR sequence of B3GalT2 (Accession No. Y15014). BLAST analysis against the GenBank htgs database finds our 5' sequence in the same BAC that contains the remainder of the gene (GenBank Accession No. AL139133). In contrast, the first 482 bp from Y15014 do not match with any genomic sequence, indicating either an alternative 5' exon or a cloning artifact.

Human Homologue of RGS8

One of the exon traps from BAC 270B19 was found to be homologous to rat RGS8 (GenBank Accession No. AB006013), a gene involved in the regulation of G-protein signaling (Saitoh *et al.*, 1997). 5'-RACE was performed to clone the complete coding sequence for human RGS8. Several alternatively spliced 5'-RACE products were analyzed to identify the one with the largest open reading frame of 198 amino acids (GenBank Accession No. AF297015). This sequence represents an alternatively spliced mRNA compared to the rat sequence and the recently submitted human RGS8 coding sequence (GenBank Accession No. AF300649), resulting in a protein with a different amino terminus (Fig. 3A). Alignment of the three cDNAs against unfinished human genomic sequences (GenBank Accession No. AL353778) shows that they are all derived from seven exons, with the rat cDNA containing exons 2 through 7, the short form of human RGS8 containing exons 3 through 7, and the long form lacking exon 3. The initiation codon (ATG) lies in exon 3 of the rat and the short form of the human RGS8 sequences. The long form of human RGS8 has its first ATG in exon 1. However, translation may initiate at the second ATG that lies in exon 2, since the first ATG has a poor Kozak sequence, containing a C rather than an A or a G at the -3 position, which generally indicates poor initiation of translation (Kozak, 1996). The alignment of the human and rat RGS8 amino acid sequences shown in Fig. 3A assumes translation initiation at the first ATG. Although we did not perform 3'-RACE, alignment of exon 7 of the coding sequence with the genomic sequence shows that the c119 cDNA sequence lies between 3.5 and 5 kb downstream of the termination codon of the RGS8 and may represent the 3'-UTR of RGS8. This would be consistent with the Northern blot data indicating that the size of the transcript is approximately 6.5 kb (Fig. 3B). Furthermore, Northern blot analysis using 16 different tissues present on MTN1 and MTN4 blots revealed expression of human RGS8 in brain only (Fig. 3B). Further analysis using a dot blot of 76 different tissues (MTE array from Clontech) showed the expression to be stronger in several subregions of the brain, particularly pons, putamen, thala-

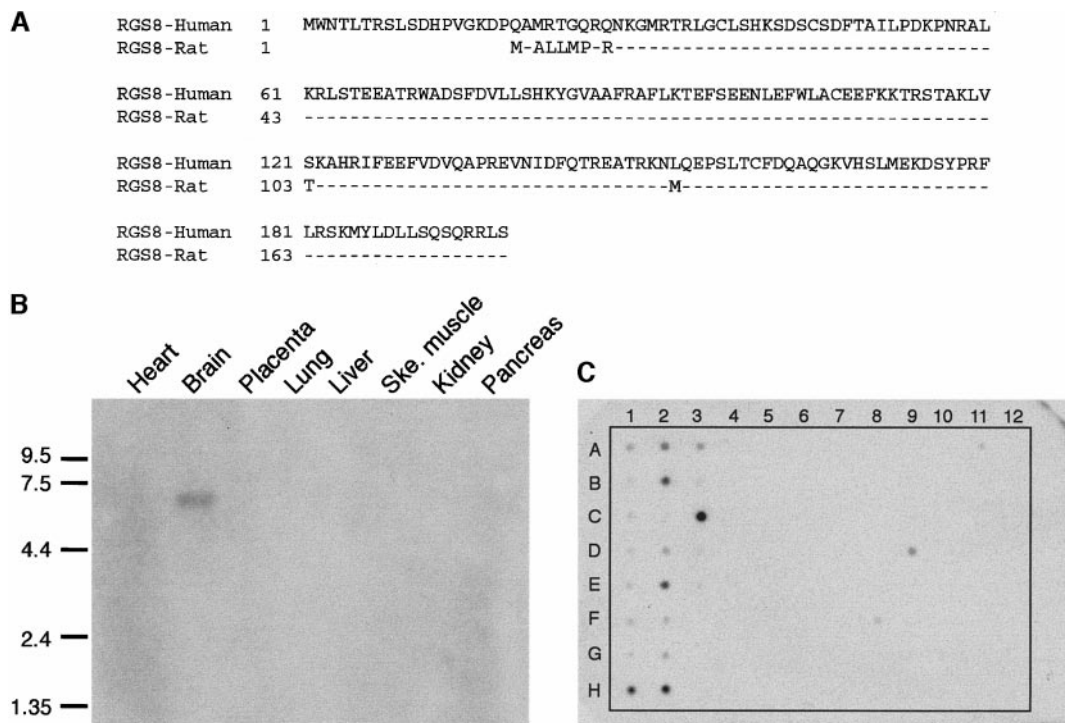


FIG. 3. Analysis of human RGS8. **(A)** Alignment of human and rat RGS8 at the amino acid level. Amino acids that differ between human and rat are shown in the rat sequence, a dashed line represents the identical residues. Amino acid numbering is given on the left side. **(B)** Northern blot analysis of the human RGS8 gene. Multiple tissue Northern blots (MTN1 and MTN4, Clontech) were hybridized with a ^{32}P -labeled cDNA probe for RGS8. Autoradiograph for MTN4 is not shown due to lack of any signal. Molecular weight standards are indicated on the left. **(C)** Expression analysis of RGS8. Multiple tissue expression array (Clontech) containing poly(A)⁺ RNA from 76 different human tissues was hybridized with the same probe as the MTN blots. Columns 1–3 contain various subregions of brain; position A11 contains fetal brain, D9 is thyroid gland, and F8 is testis.

mus, caudate nucleus, and cerebellum (Fig. 3C). In addition, a low level of expression is seen in fetal brain, testis, and thyroid gland.

Genomic Organization

The availability of genomic sequences, i.e., sequenced BAC and PAC clones as a result of the Human Genome Project, has been a tremendous help in determining the genomic organization of these transcripts. The last column of Table 2 lists the accession numbers for the genomic sequences that align with the transcripts and the number of exons in each transcript. Despite a positive signal on the Northern blots, several lines of evidence suggest that e6 and c132/e12 ESTs may not represent true transcripts. Not only do their sequences align with the genomic sequence uninterrupted (no splicing), they also lie in the introns of other transcripts (Fig. 1). Furthermore, we have assembled 7.1 kb of sequence for c132/e12 and 2.4 kb for e6, and no open reading frame is evident. The presence of e16 exons on the opposite strand of the c132/e12 sequence would explain the similar Northern signals observed using probes for both of these ESTs (Table 2).

As no genomic sequence was present in GenBank for the ge36 cDNA, we obtained the sequence of 11 exons

by directly sequencing BACs 211P21 and 112B20 using primers based on the cDNA sequence (GenBank Accession Nos. AF297016–AF297023 and AF312863). The second exon, seen in only one EST (GenBank Accession No. AI041091), is part of an *Alu* repeat and is usually spliced out.

While examining the genomic structure of the e11 cDNA, we noted the presence of a polymorphic *Alu*-Sb2 sequence (Batzer *et al.*, 1995) in the first intron of e11. This sequence is present in the BAC 120K12 sequence (GenBank Accession No. AL109865) and absent in the unfinished sequence of BAC 98H19 (GenBank Accession No. AL109956). The sequence appears to be a recent insertion as it is flanked by a perfect 16-base direct repeat (GTGAGTTGTGATTTCT), only one copy of which is present in 98H19. The more recently finished sequence of 98H19 shows no evidence of the insertion since finishing was performed to an overlap of only 100 bases. The existence of the insertion was confirmed by direct PCR of the two BACs with primers (AAAGAAGTTTTTCATCAAGACCAGGGCT and GGAAACAGGAGAAAGAGGGTCCACTATG) flanking the insertion. Subsequent analysis of 19 affected individuals from 10 prostate cancer families found 6 of 38 chromosomes containing the insertion (16%), whereas only 2 of 132 chromosomes from unmatched CEPH controls contained the insertion (1.5%). Further anal-

ysis is required to determine the statistical significance of these data. A number of recently arisen *Alu* insertion polymorphisms have been reported to show very different allele frequencies in different ethnic populations (Milewicz *et al.*, 1996; Roy *et al.*, 2000).

DISCUSSION

The critical step in gene identification following linkage analysis is to build a saturated transcript map of the implicated region and to characterize all potential transcripts for their expression profiles and putative biological functions in the cell. Such regional transcript maps also contribute toward the ultimate goal of the Human Genome Project, which is to identify all genes encoded by the human genome (Collins *et al.*, 1998). The 1q25 region flanked proximally by D1S466 and distally by D1S1642 represents approximately 6 Mb of genomic DNA and harbors the disease loci for prostate cancer (HPC1), camptodactyly–arthropathy–coxa vara–pericarditis syndrome (CACP; Bahabri *et al.*, 1998), and the hyperparathyroidism–jaw tumor syndrome. Recently, we identified mutations in the MSF gene mapped to this region (GenBank Accession No. NM-005807) in CACP patients (Marcelino *et al.*, 1999).

Here we present data on characterization of 37 potential transcripts mapping to this region. To recognize the full coding potential of this region, we used an approach consisting of expression-independent methods, such as exon-trapping and sample sequencing, in conjunction with the human gene map information from the public database. It is worthwhile to note that three transcripts, RGS8, ge36, and ge42, were initially identified by exon traps only. There is still no EST overlapping with RGS8, whereas ge36 (Unigene cluster Hs. 112017) and ge42 (No Unigene cluster) have unmapped ESTs and, therefore, could have been missed as transcripts mapping to this region by using gene map databases alone. Similarly the CACP gene was identified as mapping to this region through sample sequencing of a BAC clone. These results highlight the need to combine various methods of transcript detection to recognize the full coding potential of a given genomic region.

Our transcript characterization strategy was dictated by our primary interest in the HPC1 gene. Therefore, the first level of prioritization based on expression profiling of 37 potential transcripts led us to pursue 22 of these further. Of these, we have identified 10 novel transcripts expressed in prostate. Although none of these transcripts is exclusively or predominantly expressed in prostate, they represent potential candidates for HPC1. Since parathyroid expression was not tested in this study, all novel transcripts as well as the ESTs not characterized here should be screened for their expression in parathyroid, kidney, and other appropriate tissues to identify potential candidates for the HRPT2 syndrome. Of the 22 transcripts chosen based on Northern blot data, 14 represented 13 novel

transcripts, 2 represented artifacts of cDNA cloning, 2 represented the human homologue of rat RGS8, 1 represented the human homologue of him-4 (Manuscript in preparation), 1 represented the human homologue of mouse RGL (Sood *et al.*, 2000), and 2 represented known genes.

RGS8 is a member of a superfamily of evolutionarily conserved proteins involved in the regulation of G-protein-dependent signaling (Dohlman and Thorner, 1997). To date, more than 20 members of the RGS family (named for its function) have been described. These proteins share a sequence of approximately 120 amino acids, called the RGS domain, that is responsible for their GTPase-activating protein activity and also contain a nuclear localization signal. However, the RGS proteins localize differentially within cells as a result of differences in their amino-terminal sequences (Chatterjee and Fisher, 2000). These genes also display distinct patterns of expression (Beadling *et al.*, 1999), which in combination with differential localization within the cell is, presumably, an important means to control RGS function. The rat RGS8 gene has been shown to be a neural-tissue-specific transcript that accelerates the modulation of G-protein-coupled channels and thus might be involved in the rapid regulation of neuronal excitability upon stimulation of G-protein-coupled receptors (Saitoh *et al.*, 1997). Our analysis of the human homologue of the RGS8 gene shows it to be a neural-tissue specific transcript also. Two alternate forms of human RGS8 with different amino-terminal sequences and, presumably, different promoters might add a further level of control on their function in the neural tissues.

Of the 13 novel transcripts, we have cloned full-length cDNAs for 9 transcripts. This is evident from either the presence of an in-frame upstream termination codon or the presence of a Kozak consensus sequence and concordance between the expected sizes from the Northern blot data and the cloned length of the transcripts. Several of the proteins encoded by the novel transcripts are fairly large, e.g., greater than 900 amino acids for e11, e15, and e16.

No deductions could be made for the putative biological roles of c48/KIAA0250, e17, and e8/c24, due to the lack of any similar sequences of known function or robust sequence motifs that would suggest a function. The remaining novel transcripts demonstrated varying levels of homologies to predicted and cloned genes from mouse, rat, *Drosophila*, *C. elegans*, and yeast (Table 2). In most cases, the homology is either too weak or limited to too small a segment of the protein to make any inferences about function without further studies. However, a few of the proteins have features that suggest a potential function.

ge42 contains one or two C2H2 zinc finger domains as well as a tRNA methyltransferase domain, which suggests that it may be involved in the methylation of nucleic acids. c33/c47/KIAA0584 has substantial homology to two human proteins encoded by AF177203

(which is a carboxyl extension of KIAA1502) and AK025982. These proteins are apparently 100 and 380 amino acids shorter than c33/c47/KIAA0584 and align with it with minimal gaps over their whole length including a conserved carboxyl-terminal motif, RDEL, which is a minor variant of the endoplasmic reticulum targeting signal, which is prototypically KDEL. Starzyk *et al.* (2000) have identified AF177203 as encoding a cerebral cell adhesion molecule that supports leukocyte adhesion although it is not clear how to reconcile this function with possible sequestration in the lumen of the endoplasmic reticulum.

ge36 has significant but scattered homology to the murine protein PAL, a novel protein expressed in proliferating cells, which binds to the SH2 domain of Shc (Schmandt *et al.*, 1999). Shc in turn binds to tyrosine-phosphorylated proteins and plays a role in signaling by receptors with tyrosine kinase activity. Thus one might suppose that ge36 plays a role in the same signaling pathways.

c112 has no significant homology with any eukaryotic proteins but has 30–40% homology to various prokaryotic *N*-acetylneuraminidase lyase (EC4.1.3.3) proteins and roughly 30% homology to prokaryotic dihydrodipicolinate synthases, which extends over essentially the full length of both c112 and the prokaryotic proteins without substantial gap insertions. As both proteins have similar structures (Mirwaldt *et al.*, 1995) and use pyruvate as a substrate, it is likely that c112 will have a similar enzymatic function.

Finally, transcript identification and cloning have recently moved from wet lab to *in silico* approaches due to the availability of genomic sequences and a variety of gene-prediction programs (Stormo, 2000). Validation of such an approach is required to make sure that the entire coding potential of the human genome will be identified. Since a minimal tiling path of BAC clones representing the entire 6-Mb region is currently being sequenced at the Sanger Centre, it will be interesting to see whether all transcripts cloned during this study will be recognized by gene prediction and whether all exons of a transcript can be predicted precisely. The presence of partial protein sequences in GenBank for e11, e16, and ge42 derived from genomic sequences using gene prediction approaches suggests that such an approach is feasible but requires better programs to identify the ends of transcripts. Furthermore, our cDNA clones provide a tangible source that can be used in the laboratory to ask questions about their biological roles in the cell.

This study provides a resource of several candidate genes for HPC1 and HRPT2 syndrome. At least 10 transcripts are identified as potential candidates for prostate cancer based on their expression pattern data. In the absence of functional data on these genes, further prioritization is difficult. Therefore, evaluation of these genes as candidates requires mutational analysis, and we are currently performing mutational analysis using single-stranded conformational polymor-

phism and direct sequencing of exons using cDNA and genomic DNA, respectively, from patients from families showing linkage to the 1q25 region to identify the disease-associated mutations.

ACKNOWLEDGMENTS

The authors thank Darryl Leja for outstanding assistance with artwork. This work was supported through the National Human Genome Research Institute Intramural Program.

REFERENCES

- Ahman, A.-K., Jonsson, B.-A., Damber, J.-E., Bergh, A., Emanuelsson, M., and Gronberg, H. (2000). Low frequency of allelic imbalance at the prostate cancer susceptibility loci HPC1 and 1p36 in Swedish men with hereditary prostate cancer. *Genes Chromosomes Cancer* **29**: 292–296.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Bahabri, S. A., Suwairi, W. M., Laxer, R. M., Polinkovsky, A., Dalaan, A. A., and Warman, M. L. (1998). The camptodactyly–arthropathy–coxa vara–pericarditis syndrome: Clinical features and genetic mapping to human chromosome 1. *Arthritis Rheum.* **41**: 730–735.
- Batzer, M. A., Rubin, C. M., Hellmann-Blumberg, U., Alegria-Hartman, Leefflang, E. P., Stern, J. D., Bazan, H. A., Shaikh, T. H., Deininger, P. L., and Schmid, C. W. (1995). Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. *J. Mol. Biol.* **247**: 418–427.
- Beadling, C., Druey, K. M., Richter, G., Kehrl, J. H., and Smith, K. A. (1999). Regulators of G protein signaling exhibit distinct patterns of gene expression and target G protein specificity in human lymphocytes. *J. Immunol.* **162**: 2677–2682.
- Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J.-M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* **10**: 1001–1010.
- Berry, R., Schroeder, J. J., French, A. J., McDonnell, S. K., Peterson, B. J., Cunningham, J. M., Thibodeau, S. T., and Schaid, D. J. (2000). Evidence for a prostate cancer-susceptibility locus on chromosome 20. *Am. J. Hum. Genet.* **67**: 82–91.
- Berthon, P., Valeri, A., Cohen-Akenine, A., Drelon, E., Paiss, T., Wohr, G., Latil, A., Millasseau, P., Mellah, I., Cohen, N., Blanche, H., Bellane-Chantelot, C., Demenais, F., Teillac, P., Le Duc, A., de Petriconi, R., Hautmann, R., Chumakov, I., Bachner, L., Maitland, N. J., Lidereau, R., Vogel, W., Fournier, G., Mangin P., Cohen, D., and Cussenot, O. (1998). Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2–43. *Am. J. Hum. Genet.* **62**: 1416–1424.
- Carpten, J. D., Makalowska, I., Robbins, C. M., Scott, N., Sood, R., Connors, T. D., Bonner, T. I., Smith, J. R., Faruque, M. U., Stephan, D. A., Pinkett, H., Morgenbesser, S. D., Su, K., Graham, C., Gregory, S. G., Williams, H., McDonald, L., Baxevanis, A. D., Klinger, K. W., Landes, G. M., and Trent, J. M. (2000). A 6-Mb high-resolution physical and transcription map encompassing the hereditary prostate cancer 1 (HPC1) region. *Genomics* **64**: 1–14, doi: 10.1006/geno.1999.6051.
- Chatterjee, T. K., and Fisher, R. A. (2000). Cytoplasmic, nuclear, and Golgi localization of RGS proteins—Evidence for N-terminal and RGS domain sequences as intracellular targeting motifs. *J. Biol. Chem.* **275**: 24013–24021.
- Cher, M. L., Bova, G. S., Moore, D. H., Small, E. J., Carroll, P. R., Pin, S. S., Epstein, J. I., Isaccs, W. B., and Jensen, R. H. (1996). Genetic alterations in untreated metastases and androgen-independent

- prostate cancer detected by comparative genomic hybridization and allelotyping. *Cancer Res.* **56**: 3091–3102.
- Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L., and the members of the DOE and NIH planning groups (1998). New goals for the U.S. Human Genome Project: 1998–2003. *Science* **282**: 682–689.
- Cooney, K. A., McCarthy, J. D., Lange, E., Huang, L., Miesfeldt, S., Montie, J. E., Oesterling, J. E., Sandler, H. M., and Lange, K. (1997). Prostate cancer susceptibility locus on chromosome 1q: A confirmatory study. *J. Natl. Cancer Inst.* **89**: 955–999.
- Deloukas, P., Schuler, G. D., Gyapay, G., Beasley, E. M., Soderlund, C., Rodriguez-Tome, P., Hui, L., Matisse, T. C., McKusick, K. B., Beckman, J. S., Bentolila, S., Bihoreau, M.-T., Birren, B. B., Browne, J., Butler, A., Castle, A. B., Chiannilkulchai, N., Clee, C., Day, P. J. R., Dehejia, A., Dibling, T., Drouot, N., Duprat, S., Fizames, C., Fox, S., Gelling, S., Green, L., Harrison, P., Hocking, R., Holloway, E., Hunt, S., Keil, S., Lijnzaad, P., Louis-Dit-Sully, C., Ma, J., Mendis, A., Miller, J., Morissette, J., Muselet, D., Nusbaum, H. C., Peck, A., Rozen, S., Simon, D., Slonim, D. K., Staples, R., Stein, L. D., Stewart, E. A., Suchard, M. A., Thangarajah, T., Vega-Czarny, N., Webber, C., Wu, X., Hudson, J., Auffray, C., Nomura, N., Sikela, J. M., Polymeropoulos, M. H., James, M. R., Lander, E. S., Hudson, T. J., Myers, R. M., Cox, D. R., Weissenbach, J., Boguski, M. S., and Bentley, D. R. (1998). A physical map of 30,000 human genes. *Science* **282**: 744–746.
- Dohlman, H. G., and Thorner, J. (1997). RGS proteins and signaling by heterotrimeric G proteins—Minireview. *J. Biol. Chem.* **272**: 3871–3874.
- Dunsmuir, W. D., Edwards, S. M., Lakhani, S. R., Young, M., Corbishley, C., Kirby, R. S., Dearnaley, D. P., Dowe, A., Ardern-Jones, A., and Kelly, J. (1998). Allelic imbalance in familial and sporadic prostate cancer at the putative human prostate cancer susceptibility locus, HPC1. *Br. J. Cancer* **78**: 1430–1433.
- Gibbs, M., Stanford, J. L., McIndoe, R. A., Jarvik, G. P., Kolb, S., Goode, E. L., Chakrabarti, L., Schuster, E. F., Buckley, V. A., Miller, E. L., Brandzel, S., Hood, L., and Ostrander, E. A. (1999). Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. *Am. J. Hum. Genet.* **64**: 1087–1095.
- Hobbs, M. R., Pole, A. R., Pidwirny, G. N., Rosen, I. B., Zarbo, R. J., Coon, H., Heath, H., III, Leppert, M., and Jackson, C. E. (1999). Hyperparathyroidism–jaw tumor syndrome: The HRPT2 locus is within a 0.7-cM region on chromosome 1q. *Am. J. Hum. Genet.* **64**: 518–525.
- Kolbinger, F., Streiff, M. B., and Katopodis, A. G. (1998). Cloning of a human UDP-galactose: 2-Acetamido-2-deoxy-D-glucose 3 β -galactosyltransferase catalyzing the formation of type 1 chains. *J. Biol. Chem.* **273**: 433–440.
- Kozak, M. (1996). Interpreting cDNA sequences: Some insights from studies on translation. *Mamm. Genome* **7**: 563–574.
- Latil, A., Cussenot, O., Fournier, G., and Lidereau, R. (1997). Infrequent allelic imbalance at the major susceptibility HPC1 locus in sporadic prostate tumors. *Int. J. Cancer* **71**: 1118.
- Marcelino, J., Carpten, J. D., Suwairi, W. M., Gutierrez, O. M., Schwartz, S., Robbins, C., Sood, R., Makalowska, I., Baxevasis, A., Johnstone, B., Laxer, R. M., Zemel, L., Kim, C. A., Herd, J. K., Ihle, J., Williams, C., Johnson, M., Raman, V., Alonso, L. G., Brunoni, D., Gerstein, A., Papadopoulos, N., Bahabri, S. A., Trent, J. M., and Warman, M. L. (1999). CACP, encoding a secreted proteoglycan, is mutated in camptodactyly–arthropathy–coxa vara–pericarditis syndrome. *Nat. Genet.* **23**: 319–322.
- Milewicz, D. M., Byers, P. H., Reveille, J., Hughes, A. L., and Duvic, M. (1996). A dimorphic Alu Sb-like insertion in COL3A1 is ethnic-specific. *J. Mol. Evol.* **42**: 117–123.
- Mirwaldt, C., Korndorfer, I., and Huber, R. (1995). The crystal structure of dihydroadipicinate synthase from *Escherichia coli* at 2.5 Å resolution. *J. Mol. Biol.* **246**: 227–239.
- Nagase, T., Seki, N., Ishikawa, K., Ohira, M., Kawarabayasi, Y., Ohara, O., Tanaka, A., Kotani, H., Miyajima, N., and Nomura, N. (1996). Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201–KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain. *DNA Res.* **3**: 321–329.
- Nagase, T., Ishikawa, K., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O. (1998). Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins *in vitro*. *DNA Res.* **5**: 31–39.
- Neuhausen, S. L., Farnham, J. M., Kort, E., Tavtigian, S. V., Skolnick, M. H., and Cannon-Albright, L. A. (1999). Prostate cancer susceptibility locus HPC1 in Utah high-risk pedigrees. *Hum. Mol. Genet.* **8**: 2437–2442.
- Ohara, O., Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Seki, N., and Nomura, N. (1997). Construction and characterization of human brain cDNA libraries suitable for analysis of cDNA clones encoding relatively large proteins. *DNA Res.* **4**: 53–59.
- Rebbeck, T. R., Walker, A. H., Zeigler-Johnson, C., Weisbutg, S., Martin, A.-M., Nathanson, K. L., Wein, A. J., and Malkowicz, S. B. (2000). Association of HPC2/ELAC2 genotypes and prostate cancer. *Am. J. Hum. Genet.* **67**: 1014–1019.
- Robbins, C. M., Hsu, E., and Gillevet, P. M. (1996). Sequencing homopolymer tracts and repetitive elements. *Biotechniques* **20**: 862–868.
- Roy, A. M., Carroll, M. L., Nguyen, S. V., Salem, A.-H., Oldridge, M., Wilkie, A. O. M., Batzer, M. A., and Deininger, P. L. (2000). Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res.* **10**: 1485–1495.
- Saitoh, O., Kubo, Y., Miyatani, Y., Asano, T., and Nakata, H. (1997). RGS8 accelerates G-protein-mediated modulation of K⁺ currents. *Nature* **390**: 525–529.
- Sattler, H.-P., Rohde, V., Bonkhoff, H., Zwergel, T., and Wullich, B. (1999). Comparative genomic hybridization reveals DNA copy number gains to frequently occur in human prostate cancer. *Prostate* **39**: 79–86.
- Schmandt, R., Liu, S. K., and McGlade, C. J. (1999). Cloning and characterization of mPAL, a novel Shc SH2 domain-binding protein expressed in proliferating cells. *Oncogene* **18**: 1867–1879.
- Smith, J. R., Freije, D., Carpten, J. D., Gronberg, H., Xu, J., Isaacs, S. D., Brownstein, M. J., Bova, G. S., Guo, H., Bujnovszky, P., Nusskern, D. R., Damber, J.-E., Bergh, A., Emanuelsson, M., Kallioniemi, O. P., Walker-Daniels, J., Bailey-Wilson, J. E., Beaty, T. H., Meyers, D. A., Walsh, P. C., Collins, F. S., Trent, J. M., and Isaacs, W. B. (1996). Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. *Science* **274**: 1371–1374.
- Sood, R., Makalowska, I., Carpten, J. D., Robbins, C. M., Stephan, D. A., Connors, T. D., Morgenbesser, S. D., Su, K., Pinkett, H. W., Graham, C. L., Quesenberry, M. I., Baxevasis, A. D., Klinger, K. W., Trent, J. M., and Bonner, T. I. (2000). The human RGL (RalGDS-like) gene: Cloning, expression analysis and genomic organization. *Biochem. Biophys. Acta* **1491**: 285–288.
- Starzyk, R. M., Rosenow, C., Frye, J., Leismann, M., Rodzinski, E., Putney, S., and Tuomanen, E. I. (2000). Cerebral cell adhesion molecule: A novel leukocyte adhesion determinant on blood–brain barrier capillary endothelium. *J. Infect. Dis.* **181**: 181–187.
- Stormo, G. D. (2000). Gene-finding approaches for eukaryotes. *Genome Res.* **10**: 394–397.
- Suarez, B. K., Lin, J., Burmester, J. K., Broman, K. W., Weber, J. L., Banerjee, T. K., Goddard, K. A. B., Witte, J. S., Elston, C., and Catalona, W. J. (2000). A genome screen of multiple sibships with prostate cancer. *Am. J. Hum. Genet.* **66**: 933–944.
- Teh, B. T., Farnebo, F., Kristoffersson, U., Sundelin, B., Cardinal, J., Axelsson, R., Yap, A., Epstein, M., Heath, H., III, Cameron, D., and Larsson, C. (1996). Autosomal dominant primary hyperparathy-

- roidism and jaw tumor syndrome associated with renal hamartomas and cystic kidney disease: Linkage to 1q21-q32 and loss of the wild type allele in renal hamartomas. *J. Clin. Endocr. Metab.* **81**: 4204-4211.
- Visakorpi, T., Kallioniemi, A. H., Syvanen, A. C., Hyytinen, E. R., Karhu, R., Tammela, T., Isola, J. J., and Kallioniemi, O. P. (1995). Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer Res.* **55**: 342-347.
- Williamson, C., Cavaco, B. M., Jausch, A., Dixon, P. H., Forbes, S., Harding, B., Holtgreve-Grez, H., Schoell, B., Pereira, M. C., Font, A. P., Loureiro, M. M., Sobrinho, L. G., Santos, M. A., and Thakker, R. V. (1999). Mapping the gene causing hereditary primary hyperparathyroidism in a Portuguese kindred to chromosome 1q22-q31. *J. Bone Miner. Res.* **14**: 230-239.
- Wolfsberg, T. G., and Landsman, D. (1997). A comparison of expressed sequence tags (ESTs) to human genomic sequences. *Nucleic Acids Res.* **25**: 1626-1632.
- Xu, J., Meyers, D., Freije, D., Isaacs, S., Wiley, K., Nusskern, D., Ewing, C., Wilkens, E., Pirooska, B., Bova, G. S., Walsh, P., Isaacs, W., Schleutker, J., Matikainen, M., Tammela, T., Visakorpi, T., Kallioniemi, O. P., Ewing, C., Berry, R., Schaid, D., French, A., McDonnell, S., Schroeder, J., Blute, M., Thibodeau, S., Gronberg H., Emanuelsson, M., Damber, J.-D., Anders, B., Jonsson, B.-A., Smith, J., Bailey-Wilson, J., Carpten, J., Stephan, D., Gillanders, E., Asmundson, I., Kainu, T., Freas-Lutz, D., Baffoe-Bonnie, A., Van Aucken, A., Sood, R., Collins, F., Brownstein, M., and Trent, J. M. (1998). Evidence for a prostate cancer susceptibility locus on the X chromosome. *Nat. Genet.* **20**: 175-179.