

A Preliminary Gene Map for the Van der Woude Syndrome Critical Region Derived from 900 kb of Genomic Sequence at 1q32–q41

Brian C. Schutte,^{1,7} Bryan C. Bjork,^{1,4,7} Kevin B. Coppage,^{1,7} Margaret I. Malik,¹ Simon G. Gregory,⁵ Deborah J. Scott,⁵ Luci M. Brentzell,⁶ Yoriko Watanabe,¹ Michael J. Dixon,⁶ and Jeffrey C. Murray^{1–4,8}

Departments of ¹Pediatrics, ²Biological Sciences, and ³Preventive Medicine and Environmental Health, and ⁴Program in Genetics, University of Iowa, Iowa City, Iowa 52242 USA; ⁵Sanger Centre, Hinxton, Cambridgeshire B10 1SA, UK; and ⁶Departments of Dental Medicine and Surgery, University of Manchester, Manchester M13 9PT, UK

Van der Woude syndrome (VWS) is a common form of syndromic cleft lip and palate and accounts for ~2% of all cleft lip and palate cases. Distinguishing characteristics include cleft lip with or without cleft palate, isolated cleft palate, bilateral lip pits, hypodontia, normal intelligence, and an autosomal-dominant mode of transmission with a high degree of penetrance. Previously, the VWS locus was mapped to a 1.6-cM region in 1q32–q41 between DIS491 and DIS205, and a 4.4-Mb contig of YAC clones of this region was constructed. In the current investigation, gene-based and anonymous STSs were developed from the existing physical map and were then used to construct a contig of sequence-ready bacterial clones across the entire VWS critical region. All STSs and BAC clones were shared with the Sanger Centre, which developed a contig of PAC clones over the same region. A subset of 11 clones from both contigs was selected for high-throughput sequence analysis across the ~1.1-Mb region; all but two of these clones have been sequenced completely. Over 900 kb of genomic sequence, including the 350-kb VWS critical region, were analyzed and revealed novel polymorphisms, including an 8-kb deletion/insertion, and revealed 4 known genes, 11 novel genes, 9 putative genes, and 3 pseudogenes. The positional candidates *LAMB3*, *GOS2*, *HIRF6*, and *HSD11* were excluded as the VWS gene by mutation analysis. A preliminary gene map for the VWS critical region is as follows: CEN-VWS33-VWS34-DIS491-VWS1-VWS19-LAMB3-GOS2-VWS26-VWS25-HSD11-ADORA2BP-VWS17-VWS14-HIRF6-VWS2-VWS18-DIS205-VWS23-VWS20-VWS30-VWS31-VWS35-VWS37-VWS38-HIPP-RNASEHIP-VWS40-VWS42-VWS41-TEL. The data provided here will help lead to the identification of the VWS gene, and this study provides a model for how laboratories that have a regional interest in the human genome can contribute to the sequencing efforts of the entire human genome.

There are >300 described syndromes that have cleft lip and palate as an associated characteristic (OMIM, <http://www.ncbi.nlm.nih.gov/Omim/>). Van der Woude syndrome (VWS) is the most common form of syndromic cleft lip and palate and accounts for ~2% of all cleft lip and palate cases (OMIM no. 119300). VWS has been recognized for more than a century (Murray 1860) and was assigned its eponym following Anne Van der Woude's description of the disorder in 1954. Distinguishing characteristics include cleft lip with or without cleft palate, isolated cleft palate, bilateral lip pits, hypodontia, normal intelligence, and an autosomal-dominant mode of transmission with a high degree of penetrance (Burdick et al. 1985). VWS is distinguished from nonsyndromic cleft lip and palate by the presence of lower lip pits, which are found in most affected individuals with the disorder (Janku et al. 1980; Shprintzen et al. 1980; Burdick et al. 1985). The

unusual lip pits seen in VWS are believed to be an embryonic remnant from an early stage of development (Onofre et al. 1997), and are rarely seen in other settings. There are no other systemic, cognitive, or craniofacial anomalies to differentiate VWS from nonsyndromic forms of clefting.

Isolated clefts of the palate (CPO; secondary palate defects) are genetically and embryologically distinct from clefts that include the lip or the lip and palate together (CL/P; primary palate defects; Fraser 1955). VWS is the only single-gene form of clefting in which affected individuals within the same family commonly have either isolated cleft palate only or clefts of the lip and palate. This unique feature suggests that VWS may arise from an abnormality in a gene that disrupts a very early stage of palate development when a common factor is involved in the formation of both the primary and secondary palates.

Positional cloning of the VWS gene has progressed through genetic and physical mapping. Initially, the locus for VWS was suggested through the reporting of

⁷These authors contributed equally to this work.

⁸Corresponding author.

E-MAIL Jeff.Murray@uiowa.edu; FAX (319) 335-6970.

a patient with a large cytogenetic anomaly at 1q32–q41 by Bocian and Walker (1987) and by a suggestion of linkage to the Duffy blood group by Wienker et al. (1987). Murray et al. (1990) confirmed linkage of Van der Woude syndrome to 1q32, and subsequently, two microdeletions (Sander et al. 1994; Schutte et al. 1999) as well as individual recombinants (Schutte et al. 1996) further narrowed the region to a 1.6-cM region between the flanking markers D1S491 and D1S205. The identification of deletion mutations in three independent cases of VWS (Bocian and Walker 1987; Sander et al. 1994; Schutte et al. 1999), suggest that VWS is caused by haploinsufficiency of a gene at the VWS locus (Schutte et al. 1999). Haploinsufficiency is a common theme in autosomal-dominant clefting syndromes that include Waardenburg syndrome (OMIM no. 193500), Basal Cell Nevus syndrome (OMIM no. 109400), Rieger syndrome (OMIM no. 180500), Treacher Collins syndrome (OMIM no. 154500), and Stickler syndrome (OMIM no. 108300, 184840). In these syndromes, haploinsufficiency is evidenced by deletions and/or loss-of-function mutations (Lu-Kuo et al. 1993; Wu et al. 1993; Semina et al. 1996; Edwards et al. 1997; Wicking et al. 1997; DeStefano et al. 1998; Snead and Yates 1999). Thus, from a VWS mutation search, we expect to find a range of loss-of-function mutations in one of the positional candidates in addition to the three previously identified deletions.

The autosomal-dominant clefting syndromes described above also suggest the types of genes that would make ideal candidates for the VWS locus. Those genes encode for either transcription factors (Tassa-behji et al. 1992; Semina et al. 1996), extracellular matrix proteins (Ahmad et al. 1991) or proteins involved in signal transduction (Johnson et al. 1996). Additional candidate functions for the VWS gene can be deduced from transgenic mice whose phenotype includes an orofacial cleft. To date, ~30 knockout strains of mice exhibit some form of orofacial clefts (<http://tbase.jax.org>), and the product of those genes include growth factors in addition to transcription and signaling factors (for review, see Schutte and Murray 1999). Although genes with these functions are excellent candidates for the VWS locus, we note that not every gene involved in an autosomal-dominant clefting syndrome has such obvious developmental functions (Dixon et al. 1997).

To identify the gene responsible for VWS, we constructed a contig of bacterial clones that spans the VWS locus. STS content analysis and large-scale sequencing of this entire contig resulted in the identification of 4 known genes, 11 novel genes, 9 putative genes, and 3 pseudogenes in the 1.1-Mb region surrounding the 350-kb VWS critical region. In addition, mutation analysis excluded several positional candidates for the VWS locus.

RESULTS

Identification of Bacterial Clones from the VWS Critical Region at 1q32–q41

In a previous study (Schutte et al. 1996), we constructed a physical map of the VWS critical region composed of a 4.4-Mb contig of yeast artificial chromosomes (YACs). From that contig, we demonstrated that the VWS critical region, as defined by genetic recombinants at D1S491 and D1S205, was contained within a single 850-kb CEPH YAC clone, 785B2. To facilitate the identification of transcribed sequences in the VWS critical region, a contig of bacterial artificial chromosome (BAC) clones was constructed.

BAC clones were identified by PCR screening of the California Institute of Technology (CITB) human BAC DNA pools (B and C libraries) (Kim et al. 1996). Initially, the sequence-tagged sites (STSs) used for this screen (Table 1) were developed from three sources. The first set of STSs was obtained directly from the previous YAC physical map, which included five genetic markers—D1S245, D1S471, D1S491, D1S70, and D1S205—and five YAC end-clones—yAS9L, yAS9R, yAS10L, yAS2R, and yAS8L. When the BAC library was screened with these STSs, only a subset yielded full BAC addresses (Table 1). The STSs D1S205 and yAS8L identified the same BAC address, suggesting that they were located near each other.

The second set of STSs was developed from transcripts that were located at chromosome bands 1q32–q41. From a screen of 36 genes from 1q32–1q41 (Table 2), 4 of the genes mapped to the VWS critical region—*HSD11*, *GOS2*, Hs.179758 (*VWS31*), and *LAMB3*. One other gene, *TRAF5*, was contained on several YACs near the VWS critical region but did not amplify a product from YAC 785B2, suggesting that it was located just outside of the VWS critical region (data not shown). When the STSs from *HSD11*, *GOS2*, Hs.179758, and *LAMB3* were screened against the BAC library, all but the STS for *LAMB3* identified a full BAC address.

Finally, the third set of STSs was developed from sequences derived from the YAC clone 785B2. As described in Methods, the STSs were derived from sequences from cosmid subclones of the YAC (c53), island rescue PCR (IR2, IR6, IR7), and *Alu*-splice PCR (AS1.30, AS3.23). All of these STSs identified BAC addresses, except AS1.30 and IR2.

In total, 16 BAC clones were identified with these three sets of STSs.

Assembly of BAC Clones

Initially, the order and orientation of the BAC clones were investigated by STS content analysis with all of the STSs against all of the BAC clones (Fig. 1). The analysis indicated that the BAC clones fell into two nonoverlapping contigs. The larger contig included

Table 1. STSs Used to Construct Contigs of BAC Clones

STS	Source	GDB no.	BAC addresses from screening
D1S245	AFM224xcl	188425	221L9, 523o13
D1S471	AFM281yg1	199728	
D1S491	AFM310vb1	200285	
D1S70	D1S70 (L461)	438449	55i10
D1S205	AFM108ya3	187988	321i20
yAS9L	YAC 785B2 (left)	6530289	
yAS10L	YAC 732D12 (left)	6530231	
yAS2R	YAC 759F6 (right)	6530153	501i21, 342g16, 508k11, 447d17
yAS8L	YAC 711E10 (left)	187988	321i20, 11c7
yAS9R	YAC 785B2 (right)	6544844	5m1
LAMB3	LAMB3	9866156	
D1S3754	G0S2	9798282	501i21, 342g16
HSD11	HSD11	9866157	501i21
VWS31	HS.179758	9866159	189h6, 432o17, 672p1
AS3.23	YAC 785B2	9866161	501i21, 342g16, 508k11, 447d17
AS1.30	YAC 785B2	9866162	
VWS2 (c53)	YAC 785B2	9866163	179n7
yAS9-IR2	YAC 785B2	9798288	
yAS9-IR6	YAC 785B2	9798332	189L14, 259f11, 395g13
yAS9-IR7	YAC 785B2	9798335	189L14, 259f11
VWS18	EST 222591	9866164	21d8, 438f6
bBS77-T7	BAC 438f6	9866165	501d19
bBS78-T7	BAC 501d19	9866166	
bBS78-Sp6	BAC 501d19	9866167	

BAC clones 55i10–11c7, whereas the smaller contig included clones 432o17–189L14. At this point, all of the BAC clones and STSs used to construct these contigs were shared with the Sanger Centre. With these resources, the Sanger Centre constructed two independent contigs, ctg320 and ctg348 (<http://webace.sanger.ac.uk/cgi-bin/display?db=acedb1>). Importantly, the clones contained in these contigs were derived from a different library, the human Roswell Park Cancer Institute (RPCI) PAC library (Ioannou et al. 1994). A subset of the clones from these contigs is shown in Figure 1.

To complete the contig across this region, the CITB BAC library was rescreened with two additional STSs—*VWS18* and *bBS77-T7*. The STS for *VWS18* was derived from the 3' end of the EST 222591. The 5' end of this EST was identical to the STS *yAS8L*. Two new BAC clones were identified by the STS *VWS18*—21d8 and 438f6. The STS *bBS77-T7* was derived from the sequence of the T7 end of the BAC clone 438f6 and identified the BAC clone 501d19. STSs were produced from the sequence of each end of clone 501d19. When they were tested against the BAC clones from the two contigs, the appropriate product was amplified from BAC clones from each contig (Fig. 1), suggesting that 501d19 spanned the gap between the two contigs.

As an independent test of the order and orientation of the BAC clones, a restriction map was generated for each BAC clone and for the YAC clone 785B2 (Fig. 1). The restriction map of the BAC clones was very similar to the restriction map of the YAC clone. The few sites that were different may represent sequence

variants. One such site is the *NotI* site predicted to be about 330 kb from the centromeric end of the YAC clone 785B2. To test this hypothesis, the sequence from a subclone from this YAC clone that contained this *NotI* site was compared with the sequence from the clone 447d17 (see next section), which lacks this *NotI* site (Fig. 1). As predicted, the sequence of the YAC subclone contained the *NotI* site, GCGGCCGC, but the corresponding sequence at position 137083 of the BAC clone was GCGGCCAC. Similarly, a *MluI* polymorphism was observed in the PAC clone 434o14 at position 126196. In addition, the restriction map provided an independent test for the overlap between the BAC clone 501d19 with the distal clones 189h6, 432o17, 672p1, and 259f11. Each of these distal clones contained a *NotI* site near their centromeric ends (Fig. 1). The BAC clone 501d19 contained a corresponding *NotI* site at its distal end, providing further evidence that these clones overlap. As final confirmation of the order, orientation, and overlap of the clones in this contig, the sequence from all the BAC ends was generated (GenBank accession nos. AQ853159–AQ853198) and aligned with the sequence contigs described in the next section. Together, these results confirm that these 19 BAC clones represent a complete contig that spans ~1.1 Mb at chromosome bands 1q32–q41.

Analysis of High-Throughput Genomic Sequence

In preparation for high-throughput DNA sequence analysis at the Sanger Centre, all clones from both the BAC contig and the PAC contigs were mapped by

Table 2. Genes from 1q32–q41 Tested as Positional Candidates for VWS

Locus ^a	Mapping reference	Unigene	GenBank ^b accession no.
<i>ADORA1</i>	Townsend-Nicholson et al. (1995)	Hs.77867	P30542
<i>Abgp*</i>	Burmeister et al. (1996)	Hs.1757	P32004
<i>HSD11</i>	Tannin et al. (1991)	Hs.37012	M76665
<i>CD34</i>	Satterthwaite et al. (1992)	Hs.85289	P28906
<i>CENP-F</i>	Testa et al. (1994)	Hs.77204	U19769
<i>Chml</i>	Cremers et al. (1992)	Hs.170129	NM_001821
<i>CTSO</i>	Shi et al. (1995)	Hs.83942	U13665
<i>DYRK3</i>	Becker and Joost (1999)	Hs.38018	NM_003582
<i>ELF3</i>	Tymms et al. (1997) [#1104]	Hs.166096	AF016295
EST-A005L30	Deloukas et al. (1998)	Hs.13298	G20313
EST-A006A02 (<i>VWS31</i>)	Deloukas et al. (1998)	Hs.179758	G20640
EST-stSG4900	Deloukas et al. (1998)	Hs.29736	AA102634
EST-stSG15831	Deloukas et al. (1998)		R85605
EST-WI4396	Deloukas et al. (1998)		G02415
EST-WI9415	Deloukas et al. (1998)	Hs.79353	G07274
<i>FMOD</i>	Sztrolovics et al. (1994)	Hs.230	U05291
<i>GOS2</i>	Russell and Forsdyke (1991)	Hs.95910	M72885
<i>Ladinin*</i>	Motoki et al. (1997)	Hs.18141	T97710
<i>LAMB3</i>	Vailly et al. (1994)	Hs.75517	L25541
<i>Lgtn*</i>	Malnar-Dragojevic et al. (1997)	Hs.168670	AA603994
<i>MYOSIN</i>	Olson et al. (1990)	Hs.2830	P20428
<i>MYBPH</i>	Vaughan et al. (1993)	Hs.927	L05606
<i>NEK2</i>	Schultz et al. (1994)	Hs.153704	Z29066
<i>PCTK3</i>	Okuda et al. (1994)	Hs.2994	X66362
<i>MEKA</i>	Ding et al. (1993)	Hs.550	P20941
<i>PRELP</i>	Grover et al. (1996)	Hs.76494	U29089
<i>PROX1</i>	Zinovieva et al. (1996)	Hs.159437	U44060
<i>PTPN7</i>	Adachi et al. (1992)	Hs.35	D11327
<i>RBQ3</i>	Saijo et al. (1995)	Hs.72984	X85134
<i>SAP1</i>	Shiplely et al. (1994)	Hs.169241	M85165
<i>SSA2</i>	Chan et al. (1994)	Hs.554	M25077
<i>TAX1</i>	Kenwick et al. (1993)	Hs.2998	X68274
<i>TNNI1</i>	Corin et al. (1994)	Hs.84673	P19237
<i>TNNT2</i>	Thierfelder et al. (1994)	Hs.89749	X74819
<i>TLR5</i>	Rock et al. (1998)	Hs.114408	U88881
<i>Traf5*</i>	Nakano et al. (1997)	Hs.29736	U69108

^aGenes were mapped to 1q32–q41 by FISH and/or radiation hybrid mapping or by mapping to the mouse syntenic region (*).

^bSTSs were derived from the 3' UTR sequence found in GenBank.

FISH to verify their position at 1q32–q41 and fingerprinted to determine the degree of overlap (<http://webace.sanger.ac.uk/cgi-bin/display?db=acedb1>). On the basis of the fingerprint analysis, 11 clones from both libraries were chosen for high-throughput sequencing. The Sanger Centre has completed sequence analysis for all but two of the clones (Fig. 1; Table 3). The sequence in GenBank for the clone 55i10 (GenBank accession no. AL035408) consisted of four large noncontiguous sequence islands. To complete the sequence for this clone, PCR primers were designed at the end of the sequence islands. PCR experiments were performed with every possible combination of primer pairs to join the sequence islands. Three of the primer pairs amplified products that were 734, 308, and 771 bp in length. Because these products were relatively

small, the PCR primers were used to directly sequence the BAC clone (GenBank accession nos. AQ853199–AQ853201), thereby connecting these short sequence gaps. The complete sequence for 55i10 is 280944 bp.

To date, >900 kb of sequence has been generated over the 1.1-Mb region. The sequences for the individual clones were assembled into two sequence contigs. The larger sequence contig is 721875 bp in length and includes the entire VWS critical region. The VWS critical region was defined as the sequence between the recombinant markers D1S491 and D1S205 (Schutte et al. 1996). To determine the length of this region, the sequences for these two markers were aligned with the 720-kb contig. The nucleotide distance between these markers is 350 kb, thus defining the physical interval for the VWS critical region.

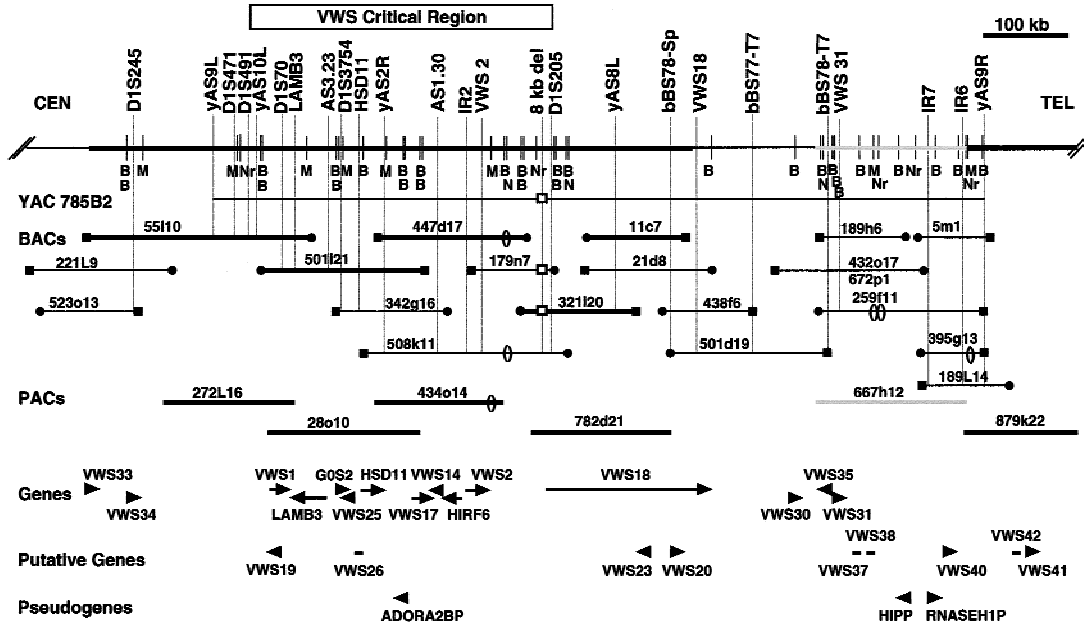


Figure 1 Physical and gene map of the VWS critical region. The VWS critical region is defined by genetic recombinants at D1S491 and D1S205 (open box at top). The CEPH YAC clone 785B2 contains both flanking markers in the original YAC contig (Schutte et al. 1996). STS content for this YAC and each of the BAC clones is indicated by long vertical lines. Restriction sites are indicated by short vertical bars and include *Bss*III (B), *Mlu*I (M), *Not*I (N), and *Nru*I (Nr). Parentheses indicate restriction sites that were absent on the indicated clone. The specific address for each BAC and PAC clone is shown. Sequencing of these clones is complete (thick line), partial (thick, shaded line), or only at their ends (thin lines). The sequence at the T7 (□) and Sp6 (●) ends for each BAC clone was determined. The presence and location of the 8-kb deletion/insertion polymorphism is indicated by an open rectangle. The order and orientation of genes, putative genes, and pseudogenes are indicated with arrowheads. For the putative genes *VWS26*, *VWS37*, *VWS38*, and *VWS42*, the direction of transcription could not be deduced (flat bar). See text for definitions of genes, putative genes, and pseudogenes.

Identification of Genes

As a first step toward constructing a complete gene map of this region, we analyzed all available sequences from this region by sequence similarity searches against the nonredundant nucleotide and peptide sequence database and nucleotide sequence EST database in GenBank. From this analysis, known genes, novel genes, putative genes, and pseudogenes were identi-

fied. In this study, known genes are defined as genomic sequences that are nearly 100% similar to a cDNA sequence derived from a gene of known function; the only sequence differences being the presence of introns and possible sequence variants. Novel genes are defined as genomic sequences that are nearly 100% similar to a cDNA or EST sequence and contain additional evidence of being a true transcript such as con-

Table 3. Summary of Genomic Sequence of BAC/PAC Clones from the VDWS Critical Region

Clone	Accession no.	Status ^a	Length ^b	Gene content ^c
BAC 55i10	AL035408	100%	280944	<i>VWS33</i> , <i>VWS34</i> , <i>VWS1</i> , <i>VWS19</i> , <i>LAMB3</i>
PAC 272L16	AL023754	C	157875	<i>VWS1</i> , <i>VWS19</i> , <i>LAMB3</i>
PAC 28o10	AL031316	C	180130	<i>VWS1</i> , <i>VWS19</i> , <i>LAMB3</i> , <i>GOS2</i> , <i>VWS25</i> , <i>HSD11</i> , <i>ADORA2BP</i>
PAC 434o14	AL022398	C	135928	<i>HSD11</i> , <i>ADORA2BP</i> , <i>VWS17</i> , <i>VWS14</i> , <i>HIRF6</i> , <i>VWS2</i>
BAC 447d17	AL022728	C	161081	<i>HSD11</i> , <i>ADORA2BP</i> , <i>VWS17</i> , <i>VWS14</i> , <i>HIRF6</i> , <i>VWS2</i>
BAC 321i20	AL035046	C	169488	<i>VWS18</i>
PAC 782d21	AL022399	C	165590	<i>VWS18</i> , <i>VWS23</i> , <i>VWS20</i>
BAC 11c7	AL022397	C	109302	<i>VWS18</i> , <i>VWS23</i> , <i>VWS20</i>
BAC 501d19		ends only	~175000	<i>VWS30</i> , <i>VWS31</i>
PAC 667h12	AL035414	>95%	~143000	<i>VWS31</i> , <i>VWS35</i> , <i>HIPP</i> , <i>VWS37</i> , <i>VWS38</i> , <i>RNASEH1P</i> , <i>VWS40</i>
PAC 879k22	AL034351	C	137679	<i>VWS42</i> , <i>VWS41</i>

^aClone statuses are either completely sequenced (C) by Sanger Centre and deposited in GenBank as a single continuous sequence, 100% completed by Sanger Centre except for some sequence derived by walking as detailed herein, sequence >95% completed by Sanger Centre, or sequenced at ends only.

^bThe clone length is based on sequence analysis, except for clone 501d19, which is based on PFGE.

^cThe gene content is based on sequence similarity searches for all clones except 501d19, which is based on STS content analysis.

sensus splice sites that flank the regions of identity and/or a polyadenylation site and polyadenylation sequence in the cDNA. Putative genes also show nearly 100% similarity to at least one EST sequence, but lack the additional evidence of being a true transcript. Pseudogenes are defined as genomic sequences that have near identity with a known gene, but generally lack introns and contain at least one mutation that disrupts the ORF of the published cDNA sequence. A total of 4 known genes, 11 novel genes, 9 putative genes, and 3 pseudogenes were identified. For each of these, we listed the genomic clone that contains the homologous sequence (Table 3). We also listed the name of the cDNA or EST sequence with the most significant similarity score, and whether any contain polyadenylation sites and sequences (Table 4). Table 4 also includes the gene clusters from human, mouse, and rat that display near sequence identity and the minimum number of exons and genomic size. The location and direction of transcription for each potential transcript is diagrammed in Figure 1.

The sequence similarity search identified four known genes—*LAMB3*, *GOS2*, *HSD11*, and *HIRF6*. Mu-

tations in *LAMB3* cause the autosomal-recessive blistering disease Herlitz junctional epidermolysis bullosa (Pulkkinen et al. 1994). *GOS2* encodes a small protein that is involved in the G0/G1 switch (Russell and Forsdyke 1991). *HSD11* encodes for an enzyme that catalyzes the interconversion of cortisol with its inactive form cortisone. There are two isozymes of this protein, and mice lacking the homolog for this isozyme displayed decreased glucocorticoid-inducible responses and were resistant to hyperglycemia caused by obesity or stress (Kotelevtsev et al. 1997). *HIRF6* is the human homolog of the mouse interferon regulatory factor 6 and belongs to a family of transcription factors that regulate the expression of interferon and interferon-stimulated genes (Grossman et al. 1996).

In addition, we identified genomic sequences with high similarity (83%–93% identity) to three other known genes, the human adenosine B2 receptor (*ADORA2B*), the HSC70-interacting protein (*HIP*), and the ribonuclease H1 (*RNASEH1*). The genomic sequence similar to the *ADORA2B* gene was identical to the previously identified pseudogene *ADORA2BP* (Jacobson et al. 1995). The *ADORA2BP* pseudogene lacked

Table 4. Genes, Putative Genes, and Pseudogenes from the VWS Critical Region

Gene	Homologous sequence	BLAST E value ^a	Unigene no. ^b			Poly(A) site	Exons ^c	Genomic size (kb) ^c
			Human	Mouse	Rat			
<i>ADORA2BP</i>	ADORA 2B pseudogene	e-142	none	none	none	N		
<i>GOS2</i>	cyclin-like	0	Hs.95910	Mm.3283	Rn.1040	Y	2	4.5
<i>HIPP</i>	HIP pseudogene	e-243		none	none	N		
<i>HSD11</i>	dehydrogenase	0	Hs.37012	Mm.28328	none	Y	6	9
<i>LAMB3</i>	laminin	0	Hs.75517	Mm.4732	none	Y	23	29
<i>RNASEH1P</i>	RNASEH1 pseudogene	e-168	none	none	none	N		
<i>VWS1</i>	rat CamPKI	e-42	R05661	W87080	A1547446	Y	>5	>30
<i>VWS2</i>	yeast ORF P40498	e-46	Hs.169207	Mm.31551	Rn.23222	Y	>12	>24
<i>VWS8</i>	HIRF6	0	Hs.11801	Mm.4179	none	Y	8	13.5
<i>VWS14</i>	EST 897026	e-60	none	none	none	Y	>1	>0.3
<i>VWS17</i>	EST 2021286	e-73	Hs.87684	Mm.32382	Rn.28176, Rn.19232	Y	>13	>26
<i>VWS18</i>	EST 222591	e-32	Hs.130270	none	none	N	5	>150
<i>VWS19</i>	EST HIBBR80	0	T09285	none	none	N	>3	>13.9
<i>VWS20</i>	EST 1505352	0	Hs.128561	none	none	N		
<i>VWS23</i>	EST 911663	e-116	AA483782	none	none	N		
<i>VWS25</i>	EST 1114940	e-112	AA601997	none	none	Y		
<i>VWS26</i>	EST 78732	e-147	AA601997	none	none	N		
<i>VWS30</i>	EST 21276	e-109	Hs.150643	none	Rn.35063	Y	>1	>0.6
<i>VWS31</i>	EST 25987	e-82	Hs.21068	none	none	Y	>2	>4.4
<i>VWS33</i>	EST 1326264	e-55	Hs.156229	none	none	Y	>1	>0.6
<i>VWS34</i>	EST 1157683	e-54	Hs.25536	none	none	Y	>4	>3.7
<i>VWS35</i>	EST 1560831	e-63	Hs.58102	none	Rn.15497	Y	>4	>2.5
<i>VWS37</i>	EST 609657	e-65	Hs.185761	none	none	N		
<i>VWS38</i>	EST 628633	e-94	None	none	none	N		
<i>VWS40</i>	EST 182501-like	e-12	Hs.99229-like	none	none	N		
<i>VWS41</i>	EST 72618	e-139	AA362845	none	none	N		
<i>VWS42</i>	EST 01691	e-114	M78101	none	none	N		

^aE value is shown as an exponent (e) with base 10.

^bFor genes that lack a Unigene number, only a single GenBank accession no. is listed, although other near identical ESTs were observed in dbEST.

^c(>) A minimum number of exons and genomic size for the gene, as the complete cDNA sequence is not known.

introns and contained numerous sequence differences that disrupted the ORF. Similarly, the sequences similar to the *HIP* and *RNASEH1P* genes also lack an ORF and lack introns indicating that they are also pseudogenes. The *HIP* pseudogene was also disrupted by two tandemly arrayed AluY repetitive elements that were flanked by a perfect 18-bp duplicated sequence.

The novel gene *VWS1* appears to be the human homolog of the rat Ca²⁺/calmodulin-dependent protein kinase 1 γ (CaMK1 γ ; B.C. Bjork and J.C. Murray, in prep.). The rat gene was isolated from brain and belongs to a family of serine/threonine protein kinases (Yokokura et al. 1997). Also of note, the *VWS2* gene is very similar to a yeast ORF of unknown function (GenPept P40498). The remaining 18 genes or putative genes display high similarity with EST sequences, including those derived from other species. However, no functional information is available for the human gene or its homologs.

Mutational Analysis of *LAMB3*, *GOS2*, *HIRF6*, and *HSD11*

To date, three independent deletion mutations were demonstrated to cause VWS (Bocian and Walker 1987; Sander et al. 1994; Schutte et al. 1999), indicating that VWS is probably caused by haploinsufficiency of the *VWS* gene. We performed mutation analysis for *LAMB3*, *GOS2*, *HIRF6*, and *HSD11*, because their complete genomic structure (Table 4) was readily determined by aligning the full-length cDNA sequence with the genomic sequence. Although these genes are poor functional candidates, our approach is to screen all positional candidates, regardless of function, as different mutations in the same gene may cause different genetic disorders (Smith et al. 1994; Biesecker 1997). From our mutation screen, we observed three sequence variants in the *LAMB3* gene, a single variant in the

HSD11 gene, and no variants in the *GOS2* nor *HIRF6* genes (Table 5). All four of these variants were observed in unaffected control samples, demonstrating that they were not etiologic mutations. These results, therefore, exclude *LAMB3*, *GOS2*, *HIRF6*, and *HSD11* as candidates for the VWS gene.

Comparison of Gene Recognition Programs

Because no etiologic mutations were found in the *LAMB3*, *GOS2*, *HIRF6*, and *HSD11* genes, additional exons must be identified and mutation analysis performed. In addition to identifying exons from full-length cDNA sequence, we performed a preliminary analysis of the genomic sequence with gene recognition programs. However, the ability of gene recognition programs to detect specific exons varies significantly (Burge and Karlin 1997). To develop an efficient strategy for identifying putative exons, the gene recognition programs contained in Genotator (Harris 1997) were compared for their ability to detect the exons in the 270-kb region that contains the *LAMB3*, *GOS2*, *HIRF6*, and *HSD11* genes (see Methods). The program Genotator performs sequence similarity searches of the GenPept and EST databases as well as gene recognition programs with Genscan (Burge and Karlin 1997), GeneFinder (Solovyev et al. 1994), GRAIL II (Xu et al. 1994), and Genie (Kulp 1996). Because the complete cDNA sequence for each of these genes tested is stored in GenBank, the GenPept search was very efficient at identifying the exons for these genes, only those exons that lacked an ORF failed to be identified with this tool (Fig. 2). Similarly, the dbEST search was also excellent for identifying exons for these genes. The exceptions were the exons located toward the 5' end of the *LAMB3* gene. This is expected, as *LAMB3* is a large gene and sequences in dbEST are biased toward

Table 5. Sequence Variants Identified from Known Genes in the VWS Critical Region

Gene	Primers (5' → 3')	Containing structure	Position	Alleles		Observed frequency ^a (%)
				major	minor	
<i>LAMB3</i>	F-CACAGTGGCAGATGAAATGC R-TCCCGTAGATGGCAAATGCT	exon 4	411 ^b	TCC(S)	TCA(S)	N.D.
<i>LAMB3</i> ^c	F-GGTGGGGCCTTCTTGATG R-TCTGCCTCCTACCTGTG	exon 17	2668 ^b	ATG(M)	TTG(L)	29
<i>LAMB3</i>	F-CCACGATGTCTGTGTCTG R-AGGTGAGATCATCAAGGCCTAGA	intron 9	+52 ^d	CCCC	CGCC	7.1
<i>HSD11</i>	F-ACCCTACTCTCCCTTGTCATT R-GAAATCCAGGATCTTCCTGC	exon 6	838 ^e	GGG(G)	GGC(G)	0.4

^aObserved frequency of minor allele based on genotyping at least 50 samples from unrelated individuals; (N.D.) not determined

^bNucleotide position in the cDNA (GenBank accession no. L25541), designating the first nucleotide of the start codon as position 1.

^cPreviously reported (Pulkkinen et al. (1995).

^dLocated 52 nucleotides from the end of exon 9.

^eOffset with respect to boundary of exon 6 of *HSD11* (GenBank accession no. M76665).

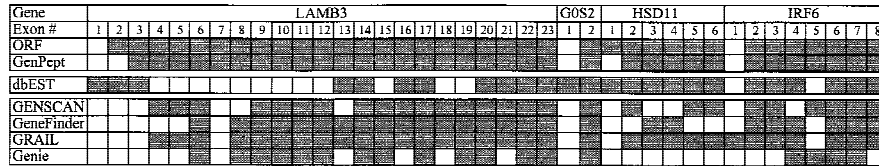


Figure 2 Comparison of gene recognition programs. Genomic sequences containing the indicated genes were analyzed for exon content with the suite of programs contained in Genotator. Exons for each gene are numbered. Those that contain part of the translated ORFs are shaded. All other shaded boxes indicate a hit with the indicated analysis program.

the 3' ends of genes because the first-strand synthesis of the cDNA clones are generally made by priming the reverse transcription reaction with an oligo-dT primer. Because the number of genes and exons in this analysis was relatively small, specific comparisons between the programs are not valid. However, consistent trends were observed. Each of these programs failed to detect the first exon of these genes and exon 7 of *LAMB3*. One of the most important criteria for each of these gene prediction programs is the presence of a coding sequence. Because the first exon of each of these genes is not translated, the failure of these programs to predict the first exon is not unexpected. Exon 7 is short (64 bp) and can be skipped without disrupting the reading frame because the subsequent exon contains an AG dinucleotide adjacent to the real splice donor site. This second potential donor site was predicted to be the splice donor site in the gene prediction programs and places the predicted gene back in frame. In addition, within the 270-kb sequence that was analyzed, whenever two or more programs predicted an exon, that sequence was an exon (Murakami and Takagi 1998). From this analysis, top priority will be given to putative exons that are predicted by multiple programs. Sequence analysis with the programs in Genotator was subsequently performed on the entire 720-kb sequence

that contains the VWS critical region. The graphical output from this analysis is available at our website (http://genetics.uiowa.edu/~bschutte/genotator_results/). In addition, a complementary set of sequence analyses was performed by the Sanger Centre and the results are available at their website (http://webace.sanger.ac.uk/cgibin/webace?db=acedb1&class=Genome_Sequence).

Identification and Verification of Novel Polymorphisms

As demonstrated with the restriction site polymorphisms, it is possible to identify sequence variants by aligning the sequences from overlapping clones that are derived from the two different libraries. Because short tandem repeats are often polymorphic, we searched the genomic sequence from this region for short tandem repeats whose copy number was different in different clones. To verify that these short tandem repeats were polymorphic, primers were designed that flanked the repeats and were then used to genotype a normal control population. As expected, all of the short tandem repeats whose sequence lengths were different in the two clones were polymorphic (Table 6). As new single nucleotide and short tandem repeat polymorphisms are discovered from this region, they are added to our web site (<http://genetics.uiowa.edu/~bschutte/polymorphisms/>).

In addition, when the sequences for the BAC clone 321i20 and PAC clone 782d21 were aligned, two regions of sequence variation were observed. Specifically, a 7922-bp sequence is absent at position 31766 of BAC clone 321i20 relative to the sequence present in PAC

Table 6. Novel STRP Markers from VWS Critical Region

Marker	GDB no.	Containing clone	Position ^a	Size (bp)	Repeat	Heterozygosity ^b
obb 55/56	9866177	55i10,28o10	221492	175	CA	0.77
obb 220/221	9866178	55i10,28o10	259697	248	AG	0.80
D1S3754	9798282	55i10,28o10	299454	233	CA	0.60
obb 62/63	9866179	55i10,272L16,28o10	354264	155	CA	0.45
D1S3753	9798272	434o14,447d17	390066	252	CCTT/TT	0.88
obb 64/65	9866180	434o14,447d17	441633	205	GT	0.55
obb 186/187	9866181	782d21	533279	212	CCTT	0.58
obb 222/223	9866182	782d21,321i20,11c7	615520	304	TA	0.68
obb 192/193	9866183	782d21,321i20,11c7	625521	247	TG	0.77
obb 190/191	9866184	782d21,11c7	651692	346	TG	0.75
obb 68/69	9866185	782d21,11c7	655252	194	GTTTTT	0.22
obb 66/67	9866186	782d21,11c7	656003	180	GTTTT	0.46
obb 188/189	9866187	782d21,11c7	684469	263	AT/AG	0.79

^aNucleotide offset from the cloning site at the T7 end of the BAC clone 55i10.

^bGenotype data from at least 50 samples from unrelated individuals.

clone 782d21 at positions 1668–9591. To prove that this deletion in BAC 321i20 was not a sequencing or cloning artifact, all clones that span this region were tested with a set of primers that flanked the deleted sequences. As shown in Figure 1, the 7922-bp sequence was present in clones BAC 508k11 and PAC 782d21 but was absent in clones YAC 785B2, BACs 179n7, and 321i20. The absence of this sequence in three independent clones derived from two different libraries demonstrates that this nearly 8-kb region represents a deletion/insertion polymorphism. Additional genotyping was performed in a normal control population and showed that the 8-kb deletion/insertion polymorphism is quite common (Y. Watanabe and B. Schutte, in prep.).

A second region of sequence variation between these two clones was observed at position 25422 in BAC clone 321i20 and position 35012 in PAC clone 782d21. The region of sequence variation was 136 bp, and the obvious difference was that the PAC clone sequence was extremely G/C rich, whereas the BAC clone sequence was not. To determine whether the variation was a sequencing artifact, we designed sequencing primers to flank this region and resequenced both clones. The new sequence was virtually identical to the PAC clone sequence, indicating that the original sequence variation was due to a sequencing or sequence assembly error in BAC 321i20.

Mouse Syntenic Region for the VWS locus

Human genes that map to 1q32 have been localized to mouse chromosome 1F and 1H (Seldin 1994), suggesting that the mouse homolog for the VWS gene could map to either 1F or 1H (Fig. 3). In this study, both *LAMB3* and *TRAF5* were mapped to the 4.4-Mb YAC contig (Schutte et al. 1996) that contains the VWS critical region. *LAMB3* is located near the proximal end of the critical region, whereas *TRAF5* is located distal to the critical region. Thus, it is probable that the VWS

gene is located between or very close to these two genes.

Because both the mouse *Lamb3* (Aberdam et al. 1994) and *Traf5* (Nakano et al. 1997) genes map to chromosome 1H, it is likely that the mouse homolog of the VWS gene also maps to mouse chromosome 1H. This information will allow us to more effectively identify positional and functional candidates from mouse. This is important because many new genes that are involved in early development are being identified from ESTs derived from embryonic tissues in mouse. No mouse phenotypes that include clefting map to this region.

DISCUSSION

The VWS locus was originally mapped by linkage analysis and microdeletions to a 1.6-cM region in 1q32–q41 between D1S491 and D1S205. In our efforts to positionally clone the VWS gene, gene-based and anonymous STSs were developed from the existing physical map and were then used to construct a contig of sequence-ready bacterial clones across the entire VWS critical region. All STSs and BAC clones were shared with the Sanger Centre, which developed a contig of PAC clones over this same region. A subset of 11 clones, derived from both the PAC and BAC contigs, were selected to represent the minimum tiling path across an ~1.1-Mb region and were the template for large-scale genomic sequencing. All but two of the clones have been sequenced completely.

We analyzed over 900 kb of genomic sequence, including the entire 350-kb VWS critical region. Our efforts revealed that novel polymorphisms and sequence similarity searches led to the construction of a preliminary transcript map encompassing the entire VWS critical region that included 4 known genes, 11 novel genes, 9 putative genes, and 3 pseudogenes. This map permitted the precise localization and transcriptional orientation of *LAMB3*, *GOS2*, and *HSD11*. These known genes were mapped previously to this region, as well as the unmapped but previously described gene *HIRF6*. Our analyses also identified a collection of potential novel genes in the VWS critical region as evidenced by the identification of genomic sequences that were nearly 100% similar to EST sequences; many of these also contained consensus splice sites and polyadenylation sites, whereas the corresponding EST sequence contained polyadenylation sequences. Additional analyses were performed with gene recognition programs on 270 kb of contiguous sequence that contains the four known genes. No single analysis was 100% specific or sensitive at identifying exons. However, by setting criteria whereby an exon must be predicted by at least two of the programs, this suite of programs successfully identified 32 of 33 exons in these four genes that contain coding sequence. These

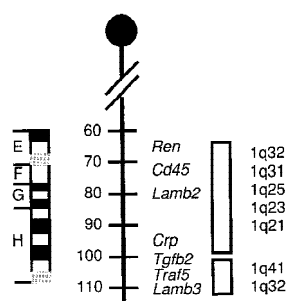


Figure 3 Mouse synteny map for human chromosome 1q32–q41. Ideogram (left) and genetic map (middle) for the distal end of mouse chromosome 1 are shown. Vertical rectangles at right represent syntenic region at the indicated human chromosome band.

results confirm the utility of these programs to predict exons and lay the foundation for identifying new exons from the region. The current gene map for the VWS critical region is as follows: CEN-VWS33-VWS34-D1S491-VWS1-VWS19-LAMB3-GOS2-VWS26-VWS25-HSD11-ADORA2BP-VWS17-VWS14-HIRF6-VWS2-VWS18-D1S205-VWS23-VWS20-VWS30-VWS31-VWS35-VWS37-VWS38-HIPP-RNASEH1P-VWS40-VWS42-VWS41-TEL. The VWS critical region was genetically defined by recombinant and deletion events at the markers D1S491 and D1S205 (Schutte et al. 1996). Thus, the genes located between these markers are positional candidates. From mutation analysis, four of these genes, *LAMB3*, *GOS2*, *HIRF6*, and *HSD11*, were excluded as candidates for the VWS gene. The contig of BAC clones presented here should greatly facilitate the identification of the gene associated with VWS.

The identification of a single gene involved in the etiology of Van der Woude syndrome will likely provide major insights into the more complex etiology of nonsyndromic cleft lip and palate. The extraordinary phenotypic overlap between VWS and nonsyndromic clefting, with only the lip pits being a differentiating factor, strongly suggests a common set of embryologic events. Because nonsyndromic clefting serves as a model for other complex traits thought to be secondary to both genetic and environmental factors, having this unique inroad into its etiology will prove beneficial in not only the specific case of clefting, but for other complex birth defects as well. Furthermore, it is possible that some of the unexplained cases of nonsyndromic clefting may, in fact, be allelic to VWS, but lack the lip pit component of the phenotype. Thus, substantial advances in the altering of recurrence risks from the 50% gene passage in the case of VWS to the lower 3%–5% recurrence risks usually given for nonsyndromic cleft lip and palate will also be of substantial benefit to families having a first affected child.

In summary, analysis of high-throughput genomic sequence is a powerful approach for developing a preliminary gene map and was greatly facilitated by the close collaboration with the Sanger Centre. Our commitment to developing a well-defined physical map at the outset and their commitment to making a sequence immediately available to the public provided us with the opportunity to investigate and analyze sequence in a real-time fashion. In addition, the interaction between our two groups was important to resolve sequence anomalies (the 8-kb deletion polymorphism and potential alignment errors) and to complete analysis of unfinished sequences. The latter will be a greater issue as the priority for the Human Genome Project shifts from finished sequence to a working-draft sequence (<http://www.nhgri.nih.gov/NEWS/news.html>). This collaborative effort provides a model for

how laboratories that have a regional interest in the human genome can contribute to the sequencing efforts of the entire human genome. The resulting sequence information, along with the descriptive approaches, including recombinant and microdeletion mapping, and close clinical characterization of the syndrome itself, provide a targeted region most suitable for gene identification.

METHODS

Patient Identification

A total of 107 individuals diagnosed with VWS and 14 individuals with popliteal pterygia syndrome (PPS) were examined at various locations, including the University of Iowa, the Philippines, and Germany. VWS and PPS families were ascertained and examined by one or more geneticists or clinical collaborators, as described previously (Schutte et al. 1999). Individuals were considered to be affected with VWS if they had one or more of the following clinical phenotypes: cleft lip, cleft palate, hypodontia, or lower-lip pits. They were considered to have PPS if they had pterygia along with any of the classical VWS phenotypic characteristics. Nearly all pedigrees had at least one individual with lip pits. We included a few pedigrees in which no family member exhibited lip pits but at least one individual had a cleft lip with or without cleft palate and at least one individual with cleft palate only. We also included families with a single affected individual that showed lip pits along with one or more features of VWS. Development and mental performance appeared to be normal in all affected individuals in this study. Ten-milliliter samples of whole blood per kilogram of body weight were obtained from the adults and 1 ml of whole blood per kilogram of body weight was obtained from each of the children in the families studied.

PCR and DNA Sequencing

DNA from these individuals was purified from whole blood (Miller et al. 1988) or blood spots (Qiagen, Germany). Approximately 40 ng of template genomic DNA were analyzed by PCR by standard conditions (GDB no. 9798291). BAC DNA was prepared as described in Identification and Characterization of BAC clones (below). Direct sequencing of BAC clones was performed by end sequencing and primer walking and carried out on an Applied Biosystems DNA Sequencing System Model 373 with fluorescently labeled dye terminators as implemented in the Taq Dye Deoxy Terminator Cycle Sequencing Kit from Applied Biosystems (Foster City, CA). Sequencing reactions included 1–2 µg of BAC DNA and 20 pmoles of primer. For end sequencing of BAC clones, standard T7 (TAATACGACTCACTATAGGG) and Sp6 (ATTTAGGTGACACTATAG) primers were used.

Development of Novel STSs from the VWS Critical Region

Novel STSs from the VWS critical region were developed by three approaches. First, STS primer pairs were designed from the 3' untranslated regions of 36 genes that were mapped previously to 1q32–q41 (Table 2). To identify genes that map to the VWS critical region, content analysis was performed with these STSs against the following DNA templates: human

genomic DNA, a hamster somatic cell line containing human chromosome 1, and YAC clone 785B2.

Second, STSs were derived from the sequence of cosmid subclones from YAC clone, yCEPH785B2, which spans the entire VWS critical region. A total of 72 cosmid clones were obtained in the sCos-1 vector by established techniques (Sambrook et al. 1989). The sequence of the T3 and T7 ends of each cosmid was determined and was then analyzed by BLAST (Altschul et al. 1990) to identify sequences with similarity to known gene sequences and/or repeats. Oligonucleotide primer pairs were designed within unique, nonrepetitive end sequence. These STSs were also tested as above to verify that they map to the VWS critical region.

Third, STSs were also obtained from the sequence of clones generated by two previously described PCR-based methods, Island Rescue PCR (Valdes et al. 1994) and *Alu*-splice PCR (Morgan et al. 1992). PCR products yielded by these methods were cloned into pBSKII, after restriction enzyme digestion, and sequenced. As above, STSs were designed to unique sequence and mapped back to the VWS critical region.

Identification and Characterization of BAC Clones

BAC clones were identified by PCR screening of the Research Genetics (Huntsville, AL) human CITB BAC library. BAC DNA was purified as recommended with either a Genome Systems KB-100 Magnum or a Qiagen QIAquick-Maxi DNA purification kit. The resulting BAC DNA was quantified via UV spectrophotometry. Following *NotI* restriction enzyme digestion, BAC clones were sized by pulsed-field gel electrophoresis (PFGE) with a CHEF DR-II apparatus (BioRad) through a 1% agarose gel in 0.5× TBE (14°C, 200 V, 5–30-sec ramp time, 14 hr). Additional rare-cutting restriction enzymes were used to generate a restriction map of the BACs. DNA from each BAC was digested singly with *BssHII*, *MluI*, and *NruI* and doubly with *MluI*–*NotI* and *NruI*–*NotI* and then analyzed by PFGE.

Assembly of Unfinished Clone Contig Sequences

The sequence for BAC clone 55i10 is presently contained in four large unfinished sequence contigs—55i10.03033, 55i10.03548, 55i10.02402, and 55i10.1090. The sequence gaps between each of these contigs was filled by direct sequence analysis of the BAC clone 55i10 with the primers derived from the ends of the unfinished sequence contigs. Prior to sequence analysis, the orientation of the unfinished contigs and the size of the gaps was determined by PCR. Each primer from a single contig was paired with all possible combinations of primers from the other contigs and used in PCR reactions with 2 ng of 55i10 as template. For these experiments, standard PCR conditions were modified to include a 3-min, 72°C extension per cycle. Product sizes and primer pairs were as follows: 734 bp for oBS1158F, 5'-GCAGCCTTACTCAATCTGAGG-3', and oBS1152R, 5'-CA-GTCAAGAAGAAATGGGCT-3'; 308 bp for oBS1153F, 5'-GGC-TACAAACCTGTACTGCA-3', and oBS1155R, 5'-GCTATG-TGCAGTGGATCACAC-3'; and 778 bp for oBS1154F, 5'-GTC-TGTGCTTCTCCATTAGC-3', and oBS1159R, 5'-GGTCA-GTTGTGTTGTGATGTC-3'. These successfully linked contigs 55i10.03033–55i10.03548, 55i10.03548–55i10.02402, and 55i10.02402–55i10.1090, respectively. The resulting sequences were aligned and sequence ambiguities resolved with the computer program Sequencher (GeneCodes, Ann Arbor, MI).

Analysis of High-Throughput Genomic DNA Sequence

Genomic sequence from the Sanger Centre was downloaded as multiple, random Unfinished sequence contigs generated from specific clones in the sequencing pipeline as soon as they became publicly available on the Sanger Centre ftp site (ftp://ftp.sanger.ac.uk/pub/human/sequences/Chr_1/unfinished_sequence/). As the initial step in analyzing genomic DNA sequence, the repetitive elements within the genomic sequence were identified and masked with RepeatMasker 2 (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>). Subsequently, BLAST (Altschul et al. 1990) analysis of these masked sequences was performed to find similarities to known gene and/or cDNA sequences in the nonredundant peptide sequence database and nucleotide sequence EST database in GenBank. Sequence similarity searches were repeated as each unfinished sequence contig was updated toward a complete Finished clone sequence contig. The finished sequences were aligned into two contigs, 721,875 and 137,679 bp (Figure 1). In addition, a 270-kb genomic sequence that contains the genes *LAMB3*, *GOS1*, *HSD11*, and *HIRF6* (nucleotides 90,000–360,000 of the 721-kb contig) was searched for putative exons with the suite of programs contained in Genotator (<http://www-hgc.lbl.gov/inf/annotation.html>). The graphical output from this analysis can be viewed at our website (http://genetics.uiowa.edu/~bschutte/genotator_results).

Mutation Analysis of Candidate Genes by SSCP

To screen for sequence variants, oligonucleotide primers were designed in the introns flanking each exon containing coding sequence for the genes *LAMB3*, *GOS2*, *HIRF6*, and *HSD11*. Each primer pair was designed to amplify a 150–200-bp PCR product for increased sensitivity in detecting SSCPs (Sheffield et al. 1993). Consequently, larger exons required multiple, overlapping primer pairs.

The exon sequences were amplified from genomic DNA from a panel of 107 unrelated individuals diagnosed with VWS and 15 unrelated individuals with PPS by standard conditions. In parallel experiments, we used the same primers to amplify PCR products from a control panel of genomic DNA from 96 CEPH grandparents and parents (Dausset et al. 1990) to determine whether detected SSCP variants were potential etiologic mutations or normal sequence polymorphisms. In addition, the inheritance of any rare variants detected in the affected panel was studied in individual affected families to determine whether they segregated with the VWS phenotype. The amplified PCR products were denatured at 95°C for 5 min and electrophoresed for 5 hr at 20 W through 0.5× MDE (FMC, Rockland, MD) acrylamide gel that was cooled with a fan. DNA bands were visualized by silver staining and inspected for potential variants. DNA fragments representing potential variants were then excised from the gel and boiled in 50 µl of ddH₂O. Ten microliters of the DNA suspension was used as template in 50-µl PCR reactions with the original primers. These PCR products were purified by extraction from 2% agarose gel with the Qiagen Gel Purification Kit (Chatsworth, CA) and sequenced in both directions. Parallel PCR products were amplified from genomic DNA of corresponding individuals possessing each potential sequence variant, purified from agarose gel, and sequenced. Sequences obtained from the variant bands were then compared with the common, normal sequence by the computer program Sequencher.

Genotyping of STRPs

Potentially polymorphic short tandem repeats were identified by aligning sequences from BAC and PAC genomic sequence by the program Sequencher (GeneCodes Corporation). PCR primers that flanked the repeats were designed with the program Primer 3.0 (<http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi>). Standard PCR conditions were used to amplify these loci from 96 CEPH grandparents and parents (Dausset et al. 1990). The products were separated on 6% acrylamide denaturing gels for 2 hr at 60 W and visualized by silver staining.

ACKNOWLEDGMENTS

We thank our collaborators at the Sanger Centre, especially the mapping, sequencing, and annotation teams, specifically Mark Vaudin and Richard Wooster for establishing the collaboration, Louise McDonald for assistance with mapping and Susan Rhodes for analysis and annotation. We acknowledge Bonnie Ludwig, Dave Spencer, Ann M. Basart, Nancy Leysens, Sheri Sekenske, Beth Pruessner, and Tom Forsha for contributory technical assistance, and Sandy Daack-Hirsch, Nancy Newkirk, and Buck Huppman for administrative support. We thank our colleagues Sue Kenwick, Martin Tymms, T.J. Yen, Walter Becker, and Margit Burmeister for sharing cDNA sequence prior to publication. We thank our many clinical colleagues (listed in Schutte et al. 1999) and their patients for contributing samples for this study. This work was supported by National Institutes of Health grants P50-DE09170 and P60-DE13076 (J.C.M. and B.C.S.), R01-DE08559 (J.C.M.), P30-HD27748 (Frank Morriss and B.C.S.), Action Research grant S/P/3261 and Colgate-Palmolive (M.J.D.). Grant support was also provided to K.B.C. by the Pediatric Scientist Development Program of the National Institute of Child Health and Human Development administered by the Association of Medical School Pediatric Department Chairman.

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

- Aberdam, D., M.F. Galliano, M.-G. Mattei, A. Pisani-Spadafora, J.P. Ortonne, and G. Meneguzzi. 1994. Assignment of mouse nicein genes to chromosomes 1 and 18. *Mamm. Genome* **5**: 229–233.
- Adachi, M., M. Sekiya, M. Isobe, Y. Kumura, Z. Ogita, Y. Hinoda, K. Imai, and A. Yachi. 1992. Molecular cloning and chromosomal mapping of a human protein-tyrosine phosphatase LC-PTP. *Biochem. Biophys. Res. Commun.* **186**: 1607–1615.
- Ahmad, N.N., L. Ala-Kokko, R.G. Knowlton, S.A. Jimenez, E.J. Weaver, J.I. Maguire, W. Tasman, and D.J. Prockop. 1991. Stop codon in the procollagen II gene (COL2A1) in a family with the Stickler syndrome (arthro-ophthalmopathy). *Proc. Natl. Acad. Sci.* **88**: 6624–6627.
- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Becker, W. and H.G. Joost. 1999. Structural and functional characteristics of Dyrk, a novel subfamily of protein kinases with dual specificity. *Prog. Nucleic Acid Res. Mol. Biol.* **62**: 1–17.
- Biesecker, L.G. 1997. Strike three for GLI3. *Nat. Genet.* **17**: 259–260.
- Bocian, M. and A.P. Walker. 1987. Lip pits and deletion 1q32-q41. *Am. J. Med. Genet.* **26**: 437–443.
- Burdick, A.B., D. Bixler, and C.L. Puckett. 1985. Genetic analysis in families with van der Woude syndrome. *J. Craniofac. Genet. Dev. Biol.* **5**: 181–208.
- Burge, C. and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**: 78–94.
- Burmeister, M., Q. Ren, G.J. Makris, D. Samson, and V. Bennett. 1996. Genes for the neuronal immunoglobulin domain cell adhesion molecules neurofascin and Nr-CAM map to mouse chromosomes 1 and 12 and homologous human chromosomes. *Mamm. Genome* **7**: 558–559.
- Chan, E.K., E.M. Tan, D.C. Ward, and A.G. Matera. 1994. Human 60-kDa SS-A/Ro ribonucleoprotein autoantigen gene (SSA2) localized to 1q31 by fluorescence in situ hybridization. *Genomics* **23**: 298–300.
- Corin, S.J., O. Juhasz, L. Zhu, P. Conley, L. Kedes, and R. Wade. 1994. Structure and expression of the human slow twitch skeletal muscle troponin I gene. *J. Biol. Chem.* **269**: 10651–10659.
- Cremers, F.P., C.M. Molloy, D.J. van de Pol, J.A. van den Hurk, I. Bach, A.H. Geurts van Kessel, and H.H. Ropers. 1992. An autosomal homologue of the choroideremia gene colocalizes with the Usher syndrome type II locus on the distal part of chromosome 1q. *Hum. Mol. Genet.* **1**: 71–75.
- Dausset, J., H. Cann, D. Cohen, M. Lathrop, J.M. Lalouel, and R. White. 1990. Centre d'étude du polymorphisme humain (CEPH): Collaborative genetic mapping of the human genome. *Genomics* **6**: 575–577.
- 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.
- DeStefano, A.L., L.A. Cupples, K.S. Arnos, J.H. Asher, Jr., C.T. Baldwin, S. Blanton, M.L. Carey, E.O. da Silva, T.B. Friedman, J. Greenberg et al. 1998. Correlation between Waardenburg syndrome phenotype and genotype in a population of individuals with identified PAX3 mutations. *Hum. Genet.* **102**: 499–506.
- Ding, C., X. Li, C.A. Griffin, E.W. Jabs, A.L. Hawkins, and M.A. Levine. 1993. The gene for human phosducin (PDC), a soluble protein that binds G-protein beta gamma dimers, maps to 1q25-q31.1. *Genomics* **18**: 457–459.
- Dixon, J., K. Hovanes, R. Shiang, and M.J. Dixon. 1997. Sequence analysis, identification of evolutionary conserved motifs and expression analysis of murine *tcof1* provide further evidence for a potential function for the gene and its human homologue, TCOF1. *Hum. Mol. Genet.* **6**: 727–737.
- Edwards, S.J., A.J. Gladwin, and M.J. Dixon. 1997. The mutational spectrum in Treacher Collins syndrome reveals a predominance of mutations that create a premature-termination codon. *Am. J. Hum. Genet.* **60**: 515–524.
- Fraser, F.C. 1955. Thoughts on the etiology of clefts of the palate and lip. *Acta Genetica* **5**: 358–369.
- Grossman, A., H.W. Mittrucker, J. Nicholl, A. Suzuki, S. Chung, L. Antonio, S. Suggs, G.R. Sutherland, D.P. Siderovski, and T.W. Mak. 1996. Cloning of human lymphocyte-specific interferon regulatory factor (hLSIRF/hIRF4) and mapping of the gene to 6p23-p25. *Genomics* **37**: 229–233.
- Grover, J., X.N. Chen, J.R. Korenberg, A.D. Recklies, and P.J. Roughley. 1996. The gene organization, chromosome location, and expression of a 55-Kda matrix protein (Prep) of human articular cartilage. *Genomics* **38**: 109–117.
- Harris, N.L. 1997. Genotator: A workbench for sequence annotation. *Genome Res.* **7**: 754–762.
- 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.
- Jacobson, M.A., R.G. Johnson, C.J. Luneau, and C.A. Salvatore. 1995. Cloning and chromosomal localization of the human A2b adenosine receptor gene (ADORA2B) and its pseudogene. *Genomics* **27**: 374–376.
- Janku, P., M. Robinow, T. Kelly, R. Bralley, A. Baynes, and M.T. Edgerton. 1980. The van der Woude syndrome in a large kindred: Variability, penetrance, genetic risks. *Am. J. Med. Genet.* **5**: 117–123.

- Johnson, R.L., A.L. Rothman, J. Xie, L.V. Goodrich, J.W. Bare, J.M. Bonifas, A.G. Quinn, R.M. Myers, D.R. Cox, E.H. Epstein, Jr., and M.P. Scott. 1996. Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science* **272**: 1668–1671.
- Kenwrick, S., M. Leversha, L. Rooke, T. Hasler, and P. Sonderegger. 1993. Localization of the human TAX-1 gene to 1q32.1: A region implicated in microcephaly and Van der Woude syndrome. *Hum. Mol. Genet.* **2**: 1461–1462.
- Kim, U.J., B.W. Birren, T. Slepak, V. Mancino, C. Boysen, H.L. Kang, M.I. Simon, and H. Shizuya. 1996. Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**: 213–218.
- Kotelevtsev, Y., M.C. Holmes, A. Burchell, P.M. Houston, D. Schmoll, P. Jamieson, R. Best, R. Brown, C.R.W. Edwards, J.R. Seckl, and J.J. Mullins. 1997. 11 β -Hydroxysteroid dehydrogenase type 1 knockout mice show attenuated glucocorticoid-inducible responses and resist hyperglycemia on obesity or stress. *Proc. Natl. Acad. Sci.* **94**: 14924–14929.
- Kulp, D., D. Haussler, M.G. Reese, and F.H. Eeckman. 1996. A generalized Hidden Markov Model for the recognition of human genes in DNA. *Intell. Systems Mol. Biol.* **4**: 1134–1142.
- Lu-Kuo, J., D.C. Ward, and R.A. Spritz. 1993. Fluorescence in situ hybridization mapping of 25 markers on distal human chromosome 2q surrounding the human Waardenburg syndrome, type I (WS1) locus (PAX3 gene). *Genomics* **16**: 173–179.
- Malnar-Dragojevic, D., Z. Trachtulec, and V. Vincek. 1997. Assignment of the mouse ligatin gene (Lgtn) to chromosome 1f by in situ hybridization. *Genomics* **40**: 192–193.
- Miller, S.A., D.D. Dykes, and H.F. Polesky. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**: 1215.
- Morgan, J.G., G.M. Dolganov, S.E. Robbins, L.M. Hinton, and M. Lovett. 1992. The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucleic Acids Res.* **20**: 5173–5179.
- Motoki, K., M. Megahed, S. LaForgia, and J. Uitto. 1997. Cloning and chromosomal mapping of mouse ladinin, a novel basement membrane zone component. *Genomics* **39**: 323–330.
- Murakami, K. and T. Takagi. 1998. Gene recognition by combination of several gene-finding programs. *Bioinformatics* **14**: 665–675.
- Murray, J.C., D.Y. Nishimura, K.H. Buetow, H.H. Ardinger, M.A. Spence, R.S. Sparkes, R.E. Falk, P.M. Falk, R.J. Gardner, E.M. Harkness et al. 1990. Linkage of an autosomal dominant clefting syndrome (Van der Woude) to loci on chromosome 1q. *Am. J. Hum. Genet.* **46**: 486–491.
- Murray, J.J. 1860. Undescribed malformation of the lower lip occurring in four members of one family. *Pract. Med. Surg.* **26**: 502–509.
- Nakano, H., M. Shindo, K. Yamada, M.C. Yoshida, S.M. Santee, C.F. Ware, N.A. Jenkins, D.J. Gilbert, H. Yagita, N.C. Copeland, and K. Okumura. 1997. Human TNF receptor-associated factor 5 (TRAF5): cDNA cloning, expression and assignment of the TRAF5 gene to chromosome 1q32. *Genomics* **42**: 26–32.
- Okuda, T., V.A. Valentine, D.N. Shapiro, and J.R. Downing. 1994. Cloning of genomic loci and chromosomal localization of the human PCTAIRE-1 and -3 protein kinase genes. *Genomics* **21**: 217–221.
- Olson, N.J., R.B. Pearson, D.S. Needleman, M.Y. Hurwitz, B.E. Kemp, and A.R. Means. 1990. Regulatory and structural motifs of chicken gizzard myosin light chain kinase. *Proc. Natl. Acad. Sci.* **87**: 2284–2288.
- Onofre, M.A., H.B. Brosco, and R. Taga. 1997. Relationship between lower-lip fistulae and cleft lip and/or palate in Von der Woude syndrome. *Cleft Palate-Craniofac. J.* **34**: 261–265.
- Pulkkinen, L., A.M. Christiano, D. Gerecke, D.W. Wagman, R.E. Burgeson, M.R. Pittelkow, and J. Uitto. 1994. A homozygous nonsense mutation in the beta 3 chain gene of laminin 5 (LAMB3) in Herlitz junctional epidermolysis bullosa. *Genomics* **24**: 357–360.
- Pulkkinen, L., J.A. McGrath, A.M. Christiano, and J. Uitto. 1995. Detection of sequence variants in the gene encoding the beta 3 chain of laminin 5 (LAMB3). *Hum. Mutat.* **6**: 77–84.
- Rock, F.L., G. Hardiman, J.C. Timans, R.A. Kastelein, and J.F. Bazan. 1998. A family of human receptors structurally related to Drosophila toll. *Proc. Natl. Acad. Sci.* **95**: 588–593.
- Russell, L. and D.R. Forsdyke. 1991. A human putative lymphocyte G0/G1 switch gene containing a CpG-rich island encodes a small basic protein with the potential to be phosphorylated. *DNA & Cell Biol.* **10**: 581–591.
- Saijo, M., Y. Sakai, T. Kishino, N. Niikawa, Y. Matsuura, K. Morino, K. Tamai, and Y. Taya. 1995. Molecular cloning of a human protein that binds to the retinoblastoma protein and chromosomal mapping. *Genomics* **27**: 511–519.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sander, A., R. Schmelzle, and J. Murray. 1994. Evidence for a microdeletion in 1q32-41 involving the gene responsible for Van der Woude syndrome. *Hum. Mol. Genet.* **3**: 575–578.
- Satterthwaite, A.B., T.C. Burn, M.M. Le Beau, and D.G. Tenen. 1992. Structure of the gene encoding CD34, a human hematopoietic stem cell antigen. *Genomics* **12**: 788–794.
- Schultz, S.J., A.M. Fry, C. Sutterlin, T. Ried, and E.A. Nigg. 1994. Cell cycle-dependent expression of Nek2, a novel human protein kinase related to the NIMA mitotic regulator of Aspergillus nidulans. *Cell Growth Differ.* **5**: 625–635.
- Schutte, B.C. and J.C. Murray. 1999. The many faces and factors of orofacial clefts. *Hum. Mol. Genet.* **8**: 1853–1859.
- Schutte, B.C., A. Sander, M. Malik, and J.C. Murray. 1996. Refinement of the Van der Woude gene location and construction of a 3.5-Mb YAC contig and STS map spanning the critical region in 1q32-q41. *Genomics* **36**: 507–514.
- Schutte, B.C., A.M. Basart, Y. Watanabe, J.J.S. Laffin, K. Coppage, B.C. Bjork, S. Daack-Hirsch, S. Patil, M.J. Dixon, and J.C. Murray. 1999. Microdeletions at chromosome bands 1q32-q41 as a cause of Van der Woude syndrome. *Am. J. Med. Genet.* **84**: 145–150.
- Seldin, M.F. 1994. Mouse chromosome 1. *Mamm. Genome* **5**: S1–S21.
- Semina, E.V., R. Reiter, N.J. Leysens, W.L. Alward, K.W. Small, N.A. Datson, J. Siegel-Bartelt, D. Bierke-Nelson, P. Bitoun, B.U. Zabel et al. 1996. Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. *Nat. Genet.* **14**: 392–399.
- Sheffield, V.C., J.S. Beck, A.E. Kwitek, D.W. Sandstrom, and E.M. Stone. 1993. The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* **16**: 325–332.
- Shi, G.P., H.A. Chapman, S.M. Bhairi, C. DeLeeuw, V.Y. Reddy, and S.J. Weiss. 1995. Molecular cloning of human cathepsin O, a novel endoprotease and homologue of rabbit OC2. *FEBS Lett.* **357**: 129–134.
- Shipley, J., D. Sheer, S. Dalton, R. Treisman, and K. Patel. 1994. Mapping of the human SAP1 (SRF accessory protein 1) gene and SAP2, a gene encoding a related protein, to chromosomal bands 1q32 and 12q23, respectively. *Genomics* **23**: 710–711.
- Shprintzen, R.J., R.B. Goldberg, and E.J. Sidoti. 1980. The penetrance and variable expression of the Van der Woude syndrome: Implications for genetic counseling. *Cleft Palate J.* **17**: 52–57.
- Smith, D.P., C. Eng, and B.A. Ponder. 1994. Mutations of the RET proto-oncogene in the multiple endocrine neoplasia type 2 syndromes and Hirschsprung disease. *J. Cell. Sci. Suppl.* **18**: 43–49.
- Snead, M.P. and J.R. Yates. 1999. Clinical and molecular genetics of Stickler syndrome. *J. Med. Genet.* **36**: 353–359.
- Solovyev, V.V., A.A. Salamov, and C.B. Lawrence. 1994. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res.* **22**: 5156–5163.
- Sztrolowics, R., X.N. Chen, J. Grover, P.J. Roughley, and J.R. Korenberg. 1994. Localization of the human fibromodulin gene

- (FMOD) to chromosome 1q32 and completion of the cDNA sequence. *Genomics* **23**: 715–717.
- Tannin, G.M., A.K. Agarwal, C. Monder, M.I. New, and P.C. White. 1991. The human gene for 11 beta-hydroxysteroid dehydrogenase. Structure, tissue distribution, and chromosomal localization. *J. Biol. Chem.* **266**: 16653–16658.
- Tassabehji, M., A.P. Read, V.E. Newton, R. Harris, R. Balling, P. Gruss, and T. Strachan. 1992. Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature* **355**: 635–636.
- Testa, J.R., J.Y. Zhou, D.W. Bell, and T.J. Yen. 1994. Chromosomal localization of the genes encoding the kinetochore proteins CENPE and CENPF to human chromosomes 4q24 → q25 and 1q32 → q41, respectively, by fluorescence in situ hybridization. *Genomics* **23**: 691–693.
- Thierfelder, L., H. Watkins, C. MacRae, R. Lamas, W. McKenna, H.P. Vosberg, J.G. Seidman, and C.E. Seidman. 1994. Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere. *Cell* **77**: 701–712.
- Townsend-Nicholson, A., E. Baker, P.R. Schofield, and G.R. Sutherland. 1995. Localization of the adenosine A1 receptor subtype gene (ADORA1) to chromosome 1q32.1. *Genomics* **26**: 423–425.
- Tymms, M.J., A.Y.N. Ng, R.S. Thomas, B.C. Schutte, J.O. Zhou, H.J. Eyre, G.R. Sutherland, A. Seth, M. Rosenberg, T. Papas, C. Debouck, and I. Kola. 1997. A novel epithelial-expressed Ets gene, Elf3 - human and murine Cdna sequences, murine genomic organization, human mapping to 1q32.2 and expression in tissues and cancer. *Oncogene* **15**: 2449–2462.
- Vailly, J., P. Szepietowski, M.G. Mattei, F. Pedoutour, R. Burgeson, J.P. Ortonne, and G. Meneguzzi. 1994. The genes for nicein/kalinin 125- and 100-kDa subunits, candidates for junctional epidermolysis bullosa, map to chromosomes 1q32 and 1q25-q31. *Genomics* **21**: 286–288.
- Valdes, J.M., D.A. Tagle, and F.S. Collins. 1994. Island rescue PCR: A rapid and efficient method for isolating transcribed sequences from yeast artificial chromosomes and cosmids. *Proc. Natl. Acad. Sci.* **91**: 5377–5381.
- Van der Woude, A. 1954. Fistula labii inferioris congenita and its association with cleft lip and palate. *Am. J. Hum. Genet.* **6**: 244–256.
- Vaughan, K.T., F.E. Weber, T. Ried, D.C. Ward, F.C. Reinach, and D.A. Fischman. 1993. Human myosin-binding protein H (MyBP-H): Complete primary sequence, genomic organization, and chromosomal localization. *Genomics* **16**: 34–40.
- Wicking, C., S. Shanley, I. Smyth, S. Gillies, K. Negus, S. Graham, G. Suthers, N. Haites, M. Edwards, B. Wainwright, and G. Chenevix-Trench. 1997. Most germ-line mutations in the nevoid basal cell carcinoma syndrome lead to a premature termination of the PATCHED protein, and no genotype-phenotype correlations are evident. *Am. J. Hum. Genet.* **60**: 21–26.
- Wienker, T.F., G. Hudek, S. Bissbort, A. Mayerova, G. Mauff, and K. Bender. 1987. Linkage studies in a pedigree with Van der Woude syndrome. *J. Med. Genet.* **24**: 160–162.
- Wu, B.L., A. Milunsky, H. Wyandt, C. Hoth, C. Baldwin, and J. Skare. 1993. In situ hybridization applied to Waardenburg syndrome. *Cytogenet. Cell Genet.* **63**: 29–32.
- Xu, Y., J.R. Einstein, R.J. Mural, M. Shah, and E.C. Uberbacher. 1994. An improved system for exon recognition and gene modeling in human DNA sequences. *Intell. Systems Mol. Biol.* **2**: 376–384.
- Yokokura, H., O. Terada, Y. Naito, and H. Hidaka. 1997. Isolation and comparison of rat cDNAs encoding Ca²⁺/calmodulin-dependent protein kinase I isoforms. *Biochem. Biophys. Acta* **1338**: 8–12.
- Zinovieva, R.D., M.K. Duncan, T.R. Johnson, R. Torres, M.H. Polymeropoulos, and S.I. Tomarev. 1996. Structure and chromosomal localization of the human homeobox gene Prox 1. *Genomics* **35**: 517–522.

Received May 27, 1999; accepted in revised form November 9, 1999.