

ARTICLE

Fine mapping of the locus for Shwachman-Diamond syndrome at 7q11, identification of shared disease haplotypes, and exclusion of *TPST1* as a candidate gene

Maja Popovic^{1,4}, Sharan Goobie^{1,4}, Jodi Morrison¹, Lynda Ellis^{2,3}, Nadia Ehtesham¹, Nicole Richards¹, Graeme Boocock^{1,4}, Peter R Durie^{2,3,5} and Johanna M Rommens^{*,1,4}

¹Program in Genetics and Genomic Biology, Research Institute, The Hospital for Sick Children, Toronto, Canada; ²Program in Integrative Biology, Research Institute, The Hospital for Sick Children, Toronto, Canada; ³Division of Gastroenterology and Nutrition, The Hospital for Sick Children, Toronto, Canada; ⁴Department of Molecular and Medical Genetics, University of Toronto, Toronto, Canada; ⁵Department of Pediatrics, University of Toronto, Toronto, Canada

Shwachman-Diamond syndrome (SDS) is an autosomal recessive disorder characterised by exocrine pancreatic dysfunction, haematological and skeletal abnormalities. We have previously defined the SDS locus as a 2.7 cM interval spanning the centromere of chromosome 7. To facilitate additional analysis of this complex and poorly characterised region, a framework of ordered genetic markers at 7p11-q11, including six newly identified, has been constructed using somatic cell, radiation hybrid and STS-content mapping. We have identified shared disease haplotypes, that recur in unrelated families of common ethnic origin, and extend across the SDS locus. Detection of ancestral and intrafamilial recombination events in patients refined the SDS locus to a 1.9 cM interval at 7q11, which contains the tyrosylprotein sulfotransferase 1 (*TPST1*) gene. Patients with SDS were screened for mutations in *TPST1* by sequencing of exons and intron–exon junctions. Two single nucleotide polymorphisms, but no disease-causing mutations, were identified. In addition, Southern blot analysis yielded no evidence of large-scale mutations, and RT–PCR analysis failed to detect alterations in expression. These results exclude *TPST1* as the causative gene for SDS. The established map of the refined SDS locus will assist in the identification and characterisation of other candidate genes for SDS. *European Journal of Human Genetics* (2002) 10, 250–258. DOI: 10.1038/sj/ejhg/5200798

Keywords: Shwachman-Diamond syndrome; 7q11; fine mapping; haplotype sharing; *TPST1* gene; exclusion

Introduction

Shwachman-Diamond syndrome (SDS, MIM 260400) is a rare multi-organ disorder with estimated incidence of 1:77 000.¹ It presents in infancy with failure to thrive and recurrent infections. The phenotypic spectrum can be broad and variable, but exocrine pancreatic dysfunction and haematological abnormalities are consistent features.²

Other common manifestations include short stature, skeletal abnormalities and liver dysfunction. Serious infections and acute myeloid leukaemia are primary concerns for morbidity and mortality in the syndrome. Segregation analysis in SDS has established autosomal recessive mode of inheritance,³ but the molecular defect remains unknown. Recently, a locus co-segregating with the disease on chromosome 7 was identified in a genome-wide scan, and recombination events defined a minimal 2.7 cM interval spanning the centromere.¹ All the families analysed supported linkage to 7p12-q11, consistent with a single locus for SDS.

To further refine the disease locus, and facilitate positional cloning of the disease gene, a linkage disequilibrium

*Correspondence: JM Rommens, Program in Genetics and Genomic Biology, Room 11-109A, Elm Wing, Annex, The Hospital for Sick Children, 555 University Ave., Toronto, ON M5G 1X8 Canada; Tel: (416) 813-7095, Fax: (416) 813-4931; E-mail: johanna@genet.sickkids.on.ca
Received 21 September 2001; revised 31 January 2002; accepted 19 February 2002

mapping approach was taken. Transmission disequilibrium test (TDT) was used as a test of association in the presence of linkage,⁴ and haplotypes were constructed to identify founder chromosome(s) and ancestral recombination events. We constructed a physical map of the disease locus using somatic cell, radiation hybrid and STS-content mapping to establish the order of published and newly characterised markers used in genetic analysis, as well as to identify candidate genes for SDS.

As the best characterised gene mapped to the critical genetic interval, tyrosylprotein sulfotransferase 1 (*TPST1*) was selected for mutation analysis, given that both its function and expression pattern are consistent with multi-organ defects observed in SDS. It is widely expressed, and codes for a Golgi transmembrane protein involved in posttranslational modification of secreted and membrane-bound proteins.⁵ Involvement of *TPST1* in the syndrome was assessed by exon sequencing, Southern blotting and expression analysis.

Materials and methods

Patients

Most of the families with SDS included in the study have been described elsewhere.^{1,2,6} Additional families were obtained through ongoing recruitment. Diagnosis of SDS was based on documented evidence of exocrine pancreatic and bone marrow dysfunction, with the latter most commonly involving chronic neutropenia.² Consent was obtained from all participating families, and procedural approval was obtained from the human subjects review board of The Hospital for Sick Children (HSC). Genomic DNA was isolated from leukocytes or EBV transformed lymphoblasts.⁷

Genotyping and genomic library screening

Primer sequences were obtained from the Genome Database (<http://gdbwww.gdb.org/>) or designed using Primer3 software.⁸ Microsatellite marker alleles were amplified by PCR, resolved on a 6% sequencing gel, transferred to Hybond-N+ membrane and visualised by hybridisation to one of the PCR primers end-labelled with [α -³²P] dCTP.⁹ SNP markers were typed by allele-specific PCR.⁹ New polymorphic markers derived from CEPH YAC 763g2 were identified by hybridisation of (CA)₁₅ oligonucleotide to 800 YAC subclones,¹¹ followed by sequencing.

Library screening of the chromosome 7-specific YAC library¹² and the RPCI-11 BAC library¹³ was performed by PCR and oligonucleotide hybridisation, respectively.

Mutation analysis

DNA from anonymous unrelated individuals of Canadian origin was obtained from The Centre for Applied Genomics, HSC. Primers for amplification and sequencing were selected by the Primer3 software. Sequencing was performed using

radiolabelled terminators (Amersham Pharmacia Biotech), and included at least 19 nucleotides of intron sequence at the intron-exon junctions. Sequence variants were characterised by bi-directional PCR amplification of specific alleles (Bi-PASA).¹⁴ For Southern blot analysis, genomic DNA was digested separately with *HincII* and *BglI*. The three probes used for hybridisation were amplified from genomic DNA or cDNA, and correspond to partial exon I (nucleotides 164–859 of *TPST1* mRNA), complete exons II–IV (nucleotides 928–1223), and partial exon V (nucleotides 1282–1754). For RT–PCR analysis, total RNA was isolated from EBV transformed lymphoblasts,¹⁵ and reverse transcribed with SuperScriptTMII (Life Technologies) using random hexamers. The BLAST algorithm was used for database search and alignment.¹⁶

Results

Genetic analyses

The linkage disequilibrium study included 34 families of diverse ethnic origin from Canada, USA, Central and South America, Europe and Australia. Patients, parents and any unaffected siblings were genotyped at polymorphic marker loci extending across the SDS locus. Alleles at 19 marker loci (from *D7S1485* to *D7S482* in Figure 1, excluding *RL15*, *RL14*, *B236I1* and *D7S2503*) were evaluated for association with disease using TDT. Transmissions from heterozygous parents to affected offspring of 34 parent–child trios were included in the test. After correction for multiple testing, none of the marker alleles showed significant excess transmission (positive association) or non-transmission (negative association) (data not shown).

Linkage disequilibrium, however, was detected by haplotype analysis. Haplotypes were constructed on disease chromosomes (chromosomes of affected individuals) and normal chromosomes (parental chromosomes not transmitted to affected offspring). Their comparison identified six haplotypes (I–VI in Figure 1), each of which occurred on two disease chromosomes of common ethnic origin and on none of the normal chromosomes. Notably, such extensive haplotype sharing was not detected among normal chromosomes. These observations suggest existence of multiple founder chromosomes in SDS.

To detect historical recombination events that may have occurred on these chromosomal backgrounds, haplotype analysis was extended to other available SDS families. Ancestral recombinants associated with disease haplotype II were observed in families SW93 and SW18, and those associated with haplotype IV were detected in family SW69 (Figure 1). The recombination events reduce the most likely location for the disease defect to an interval that is flanked by markers *BS126* and *D7S502*, and is spanned by all six SDS haplotypes.

Refinement of the disease locus was supported and extended by observation of an intrafamilial recombination

SDS chromosome		I		II				III		IV			V		VI	
		SW17-1-M	SW67-1-P	SW108-1-P	SW118-1-P	SW93-1-M	SW18-1-M	SW80-1-P	SW41-1-P	SW97-1-P	SW10-3-M	SW69-2-M	SW19-1-P	SW73-1-P	SW42-1-P	SW51-1-M
cM	Marker															
73.84	D7S506	7	5	9	9	7	9	8	10	4	3	10	9	9	1	9
74.38	D7S2552	4	2	6	6	3	6	5	6	7	6	4	5	5	5	8
74.38	D7S1485	5	5	3	3	5	5	1	1	5	5	1	3	3	5	6
-	WIAF-2179	1	1	1	1	2	2	1	1	1	1	2	1	1	1	1
75.98	D7S499	8	8	7	7	3	4	7	7	[7]	[8]	5	8	5	6	5
75.98	D7S659	5	5	7	7	7	1	8	8	4	4	5	4	4	4	5
76.71	D7S473	9	9	[6]	[7]	2	2	[7]	[5]	6	6	7	3	3	4	4
76.71	D7S494	3	3	3	3	3	3	1	1	5	5	5	4	4	5	5
-	D7S1480	5	5	4	4	4	4	4	4	2	2	2	2	2	2	2
-	BS126	5	5	9	9	9	9	9	9	6	6	1	6	6	1	1
76.71	D7S2429	2	2	2	2	2	2	5	5	3	3	3	7	7	2	2
-	RL11	11	11	[6]	[7]	[7]	[6]	[11]	[10]	10	10	10	10	10	9	9
-	RL12	5	5	3	3	3	3	6	6	5	5	5	5	5	5	5
-	WIAF-183	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
77.91	D7S2530	5	5	2	2	2	2	5	5	5	5	5	5	5	5	5
77.91	D7S2512	2	2	9	9	9	9	3	3	2	2	2	2	2	7	7
-	RL15	5	5	2	2	2	2	5	5	5	5	5	5	5	3	3
77.91	D7S520	2	2	4	4	4	4	2	2	3	3	3	3	3	3	3
-	RL14	3	3	7	7	7	7	3	3	3	3	3	3	3	7	7
77.91	D7S2549	2	2	1	1	1	1	1	1	1	1	1	2	2	1	1
-	B23611	4	4	4	4	4	4	4	4	[3]	[3]	[4]	4	4	5	5
78.65	D7S663	8	8	[9]	[8]	[9]	[8]	1	1	5	5	5	4	4	2	2
78.65	D7S502	11	11	8	8	8	8	5	5	7	7	9	3	7	8	8
78.65	D7S2503	5	5	2	2	11	7	7	7	7	2	7	6	4	9	2
78.65	D7S482	5	5	6	6	6	6	1	2	2	2	5	2	2	5	5
78.65	D7S2483	6	6	2	5	2	3	2	5	2	2	3	5	5	5	2
83.99	D7S1816	6	4	5	3	5	7	5	6	5	6	6	4	5	4	2
Ethnic origin		G	G	D	D	D	I	D	D	I	I	E	FC	FC	E	E
Country of origin		US	US	Ne	Ne	Ca	US	Ne	US	It	Ca	UK	Ca	Ca	UK	US

Figure 1 Shared disease haplotypes in SDS. Each of the six extended haplotypes (I–VI, outlined) was observed to co-segregate with the disease in two unrelated families of common ethnic origin (G, German; D, Dutch; I, Italian; FC, French Canadian; E, English). The families are from the US, the Netherlands (Ne), Italy (It), Canada (Ca) and the UK. All haplotypes span a smaller interval defined by recombination events on haplotypes II and IV. Chromosome designation refers to the SDS pedigree, the affected individual, and the parental origin (M, maternal; P, paternal). Alleles assumed to have arisen by microsatellite mutation (in brackets) differ in size by one or two repeat units. All markers involve dinucleotide or tetranucleotide repeats, with the exception of two SNPs, *WIAF-2179* and *WIAF-183* (Table 1).⁴⁴ Marker order is based on the sex-averaged genetic map (cM),³¹ the Stanford G3 radiation hybrid map, v 2.0,¹⁸ and the physical map in Figure 3. Markers *D7S473* and *D7S494* were resolved by recombinations in two families (data not shown). Relative order of markers *D7S1485* vs. *WIAF-2179*, and *D7S494* vs. *D7S1480* remains undetermined, but has no effect on either the extent of haplotype sharing, or the interpretation of recombination events.

event in SDS family SW158, where the phase of the recombinant maternal chromosome was confirmed by genotyping first degree maternal relatives in an extended pedigree (Figure 2). The region of concordance between the affected and an unaffected sibling could be excluded from involvement in disease, thereby defining a new boundary for the SDS locus at marker *D7S2429*. Taken together, the ancestral and the intrafamilial recombinants position the SDS gene in the minimal 1.9 cM interval between *D7S2429* and *D7S502* (summarised in Figure 3).

Detection of the critical recombination events described above was facilitated by characterisation of six new polymorphic markers (Table 1), with at least six alleles observed for each marker in the SDS pedigrees. Markers *BS126* and *B23611* were

identified using the publicly available sequence of RPCI-11 BAC clones R-211B24 (TIGR: <http://www.tigr.org/>) and R-458F8 (GenBank¹⁷ at NCBI), respectively. Markers *RL11*, *RL12*, *RL14* and *RL15* were derived from CEPH YAC 763g2, selected for its large size and unambiguous localisation to 7q11.2 by fluorescence *in situ* hybridisation (YAC/BAC FISH Mapping Resource at MPIMG: <http://www.mpimg-berlin-dahlem.mpg.de/~cytogen/>).

Physical mapping

In order to establish the relative order of published genetic markers, as well as those identified in this study, we used several complementary mapping methods. Initially, radiation hybrid mapping was performed, using the Stanford G3 panel.¹⁸ Several markers across the SDS locus (*D7S499*,

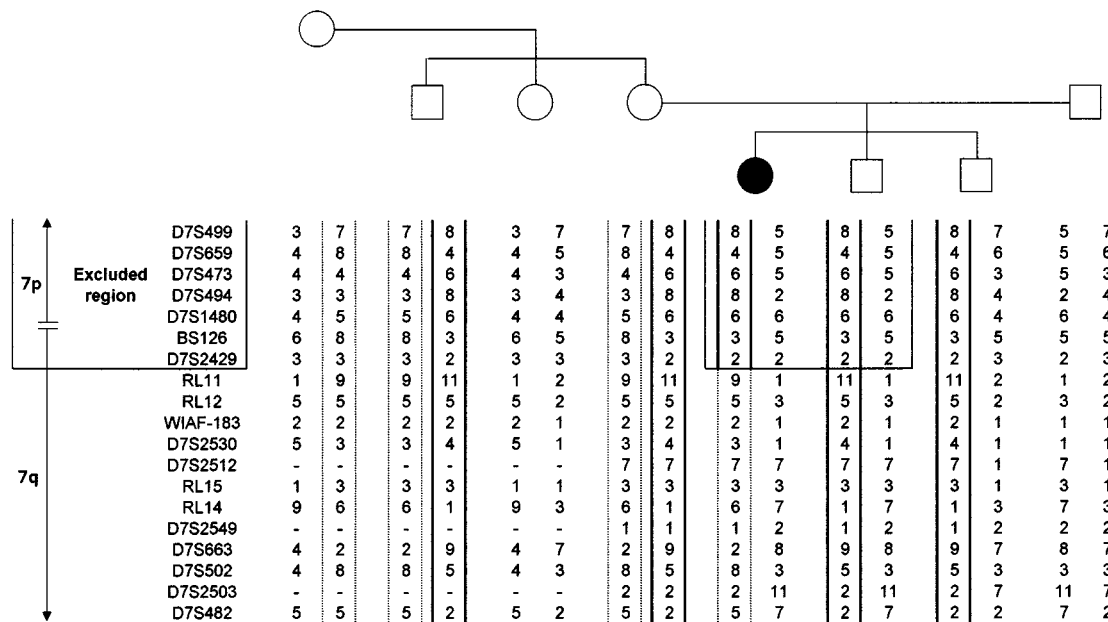


Figure 2 Recombination mapping in family SW158. The exclusion of the short arm and the centromeric region of chromosome 7 from disease involvement is based on concordant and discordant chromosomal regions in the affected and an unaffected sibling, determined by a recombination event on the maternal chromosome. The phase of the maternal chromosomes was confirmed by genotyping maternal relatives (dashes indicate undetermined genotype at uninformative markers).

D7S659, *D7S2429*, *WIAF-183*, *D7S2549*, *D7S663*, *D7S502*, *D7S2503* and *D7S482*) had already been mapped at the Stanford Human Genome Center (SHGC),¹⁸ or by others (*D7S1480*).¹⁹ An additional 11 markers were typed twice on the panel, and data were submitted to the SHGC RH server for analysis. All markers were placed in high-confidence (1000:1) bins on the Stanford G3 RH map, v 2.0¹⁸ (Figure 3) with lod scores greater than 7. Markers positioned in the same bin could not be ordered with high confidence, but were resolved by observation of intrafamilial recombination events (see Figure 1 legend), and by mapping on a YAC/BAC clone contig constructed across the 1.9 cM SDS locus.

The contig, shown in Figure 3, was established by STS-content mapping of YAC clones retrieved from CEPH and Washington University libraries based on information in published physical maps.^{20,21} In addition, the chromosome 7-specific YAC library and the RPCI-11 BAC library were screened for marker-positive clones. At later stages, physical mapping incorporated STSs developed from available BAC-end sequence (TIGR Database; Table 1), as well as EST-derived STSs predicted to map to the SDS locus on GeneMap'99.²²

Based on integrated maps,^{21,23,24} it was evident that the 2.7 cM SDS locus spanned the centromere of chromosome 7. As the interval was saturated with genetic markers, the question arose as to which markers flank the centromere on the p and the q arms. This was of special interest in determining the chromosomal location of the critical recombination event in family SW158. To orient markers

with respect to the alphoid domains of the centromere, two distinct chromosome 7-specific centromeric repeats,²⁵ *D7Z1* (Table 1) and *D7Z2*,²⁶ were radiation hybrid mapped on the Stanford G3 panel and correlated with mapping results for the genetic markers (Figure 3). Chromosome arm assignment was directly confirmed by mapping markers and centromeric repeats on somatic cell hybrids RuRag 14-4-7-44 and RuRag 6-20-12 containing the short and the long arm of human chromosome 7, respectively²⁷ (Figure 3). Together, the above findings position the refined 1.9 cM disease locus on the long arm of chromosome 7.

The refined SDS locus exceeds 1.9 Mb, the size predicted by the average 1 cM/Mb, based on radiation hybrid mapping and the constructed clone contig (see Discussion). The large physical size, as well as the pericentromeric location of the disease locus present difficulties for identification and characterisation of candidate genes. GeneMap'99 positions 28 transcripts within the critical genetic interval. To date, 13 ESTs and three known genes, *TPST1*, *PMS2L4*, and *ZFD25*, have been confirmed by us to map to the disease locus (Figure 3).

Mutation analysis of *TPST1*

In contrast to *PMS2L4* and *ZFD25*, *TPST1* is known to encode a functional protein product,⁵ and thus represents a true candidate gene for SDS. To facilitate mutation screening of *TPST1*, we established its genomic structure (Table 2). *TPST1* mRNA sequence (AF038009) was aligned to the working draft

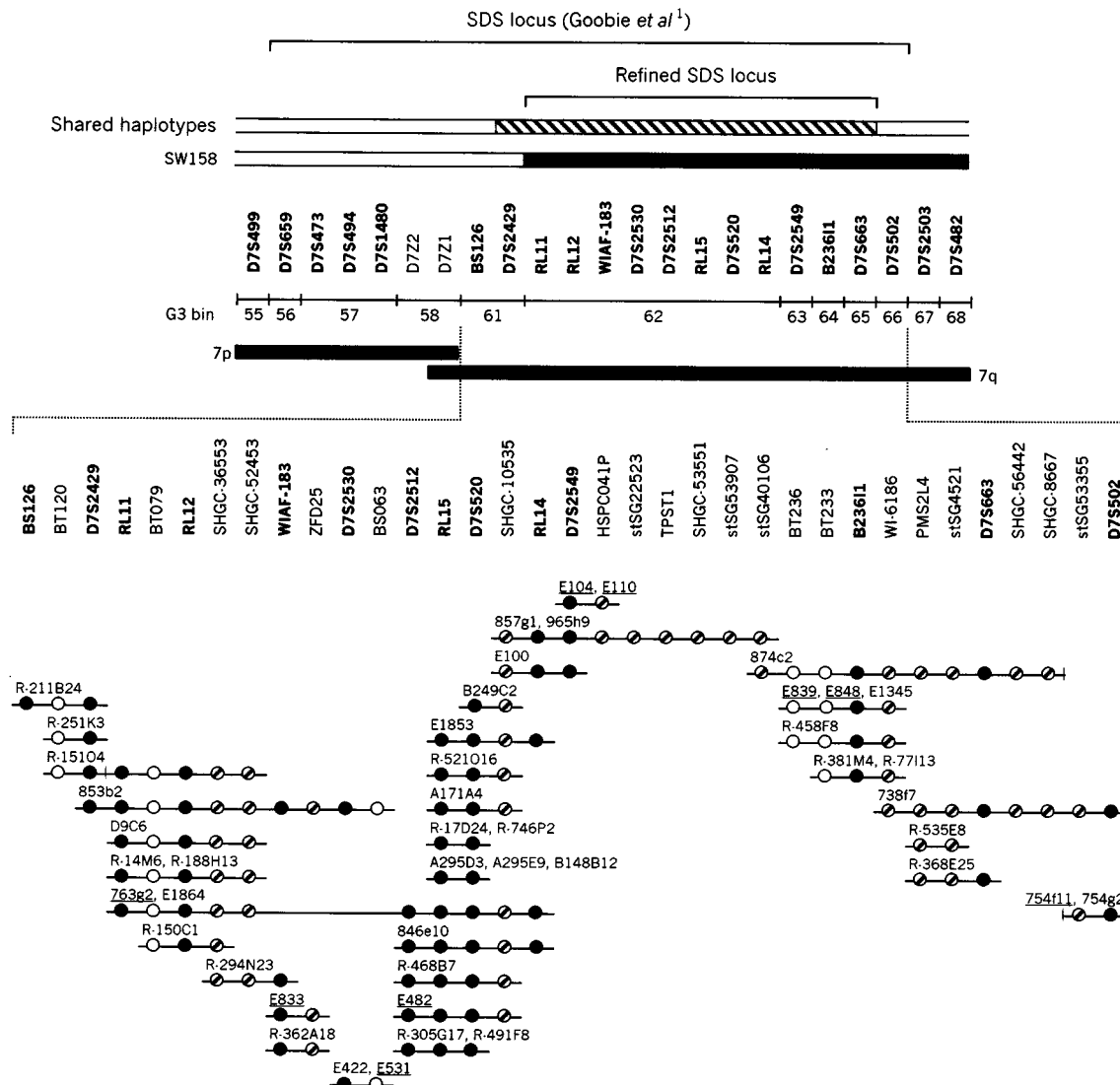


Figure 3 Refinement and physical map of the SDS locus. The SDS locus refinement is based on ancestral recombinants observed on shared disease haplotypes (minimal interval is shown as hatched bar), as well as on the pattern of concordance (open bar) and discordance (solid bar) of the affected and an unaffected sibling in SDS family SW158. Marker order is based on radiation hybrid mapping (G3 bin) and STS-content mapping. The clone contig across the disease locus includes YAC clones from CEPH, Washington University and chromosome 7-specific libraries, and BAC clones from RPCI-11 library. In addition to genetic markers (in bold; amplification shown as solid circles), the physical map incorporates BAC-end STSs (open circles) and STSs derived from known genes and ESTs (hatched circles). The refined SDS locus is at 7q11 based on mapping of centromeric repeats *D7Z1* and *D7Z2*, and on characterisation of somatic cell hybrids RuRag 14-4-7-44 (7p) and RuRag 6-20-12 (7q). In addition, many of the YAC clones in the contig (underlined) have been analysed by FISH and localised to 7q11.2 (Kunz *et al*; ⁴⁵ Integrated Chromosome 7 Database: <http://www.genet.sickkids.on.ca/chromosome7/>; YAC/BAC FISH Mapping Resource at MPIMG). For simplicity, a large number of BAC clones are omitted from the contig, as well as YAC clones with apparent deletions, with the exception of 763g2, and E1864; the former is shown as a reference point for polymorphic markers derived from this clone.

To ensure that haplotype construction and interpretation of recombination events was correct, knowledge of marker order was essential. The resolution of genetic maps across the SDS locus is low.^{30,31} Furthermore, discrepancies were noted between genetic and physical maps.^{20,21} Survey of genomic

sequence available in GenBank for this region revealed poor coverage with high degree of discontinuity, insufficient to resolve the inconsistencies. This is not surprising given the complex organisation of pericentromeric regions, including that of chromosome 7.³² The associated difficulties in

Table 2 Genomic structure of the *TPST1* gene

Exon	Location in mRNA ^a	Length (bp)	Sequence at intron–exon junctions ^b	Intron	Length (kb)
I	1–926	926	GTAGACTGT...TCTGTCAA A gtgagtaga	1	45.2
II	926–1125	199	tggtttcag A GTGGAGAG...ACTCGAAGG G taagtgag	2	75.8
III	1126–1176	51	ttttccag G TCTATAAG...AAACCACAG g tactgtgt	3	4.3
IV	1177–1223	47	cttcaacag A CTGAGCAA...ATACATGAG g tgagggtt	4	3.0
V	1224–1768	545	tcgccctag G AAAGATTG...TCAATGTGA		

^aRefers to GenBank accession number AF038009. ^bConsensus sequence at the 3' and 5' splice sites is shown in bold.

Table 3 Primers used for amplification and sequencing of *TPST1* exons, and for characterisation of detected sequence variants

Amplification target	Method	Primer sequence (5'-3') ^a	Polarity
Exon I	PCR	tccagaaaaggcttctcagtt ttccttctctactagatgttg	F R
Exon II	PCR	gtgggttaagcatgatttc cgtattcagatcccaattc	F R
Exon III	PCR	aatgcaagagccaatctagt acctgttgcttaactcac	F R
Exon IV	PCR	gtgggaacagaaagacct cttctcatgctgaccaact	F R
Exon V	PCR	tgcttctctgaaatcacatt gaaggaagagacgaagacac	F R
Exons II–IV	RT-PCR	GTGGAGAGATCTACAGACCA CTCATGTATGGAAGAGGCT	F R
1126–19(G→A)	Bi-PASA ^b	cttcaatgcaagagccaatct <i>ggggccgggggggaaaaacaaagcaggt</i> <i>cggggccgggggtgccttaactcacg</i> atgtggcagtcaggggtgt	F F R R
1599(T→G)	Bi-PASA ^b	TCCCAACTTCTCAGAAACAATACAA GGGGGCCGGGACGTGTTCTACAAAACC GGCGGGGGCCTTGTCTGTACATGTTCTAAT CCCCTGAGAGGATGGAGGT	F F R R

^aIntron sequence is indicated in lower case, and exon sequence in upper case. ^bAllele-specific primers in the Bi-PASA assay contain a noncomplementary 5' tail, shown in italics; the ultimate nucleotide, depicted in bold, corresponds to the sequence variant.

mapping and sequence assembly³³ emphasise the utility of our physical map for fine mapping studies, and underscore the value of the newly identified genetic markers.

The predicted size of the critical 1.9 cM interval, based on the Stanford G3 RH map, v 2.0, is 3.3 Mb, and yields 0.58 cM/Mb ratio of genetic and physical distance. This deviation from the average 1 cM/Mb appears to be due to reduced male recombination rate.³¹ Given the established correlation between the local recombination rate and the extent of linkage disequilibrium,^{34,35} the former may be directly related to extensive disease haplotype sharing, observed over large physical distances (based on 0.58 cM/Mb, 4.6 Mb for haplotypes V and VI, 7.3 Mb for haplotypes III and IV, and 16.5 Mb for haplotype I; Figure 1).

Since SDS is associated with a broad spectrum of clinical manifestations, it is difficult to predict the identity of the defective gene based on its functional features. Therefore, all genes in the critical genetic interval should be considered as candidates. Assessment of candidate genes is further complicated by the pericentromeric location of the refined SDS locus, where repetitive elements, including gene-related

sequences, are abundant.³² This is well illustrated by the *HSPC041* gene.³⁶ Based on sequence alignment, it is located on chromosome 8, with highly related pseudogenes found on chromosomes 7 (Figure 3), 15, and 17. Of the three known genes mapped to the SDS locus, two (*PMS2L4*³⁷ and *ZFD25*³⁸; Figure 3) belong to gene families with multiple members mapped at 7q11, which is reminiscent of duplicated sequences associated with pericentromeric regions. In the absence of functional studies, it remains unclear whether these represent pseudogenes, or should be considered as candidate genes for SDS.

TPST1 is the best-characterised gene mapped to the SDS locus. Both publicly available^{22,39} and our own mapping data position *TPST1* unambiguously within the SDS critical region (Figure 3), with no evidence of duplication locally or elsewhere in the genome. Broad tissue distribution of *TPST1* mRNA (Northern blot analysis in Ouyang *et al*⁵) is consistent with multi-organ defects observed in SDS. This includes expression in the pancreas and liver,⁵ in neutrophils,⁴⁰ and in the bone marrow (RT-PCR analysis by M Popovic, unpublished data), all of which are affected in SDS. Finally,

functional implications of a defect in *TPST1* are not inconsistent with the disease phenotype. Tyrosine sulfation of secreted and membrane-bound proteins is known to mediate protein–protein interactions involved in inflammatory and immune responses, such as leukocyte adhesion and chemokine signalling.⁴¹ This type of post-translational modification may be involved in signalling pathways that are of direct relevance to SDS. Although tyrosine sulfation has not been shown to be a dynamic process like phosphorylation, localisation of *TPST1* and arylsulfatase E (*ARSE*) in the same subcellular compartment does support this possibility.⁴² Interestingly, mutations in *ARSE* gene lead to abnormalities in the skeletal system and short stature,⁴² which are components of the SDS phenotype.

Mutation analysis presented here excludes *TPST1* as the candidate gene for SDS. Sequence analysis of exons and intron–exon junctions did not detect variants associated exclusively with disease. The identified sequence variants appear to co-segregate with SDS in affected families; however, they are positioned in the non-coding sequence, and also occur in controls. Recently, the same variants were observed in healthy Japanese individuals screened for single nucleotide polymorphisms.⁴³ We also found no evidence for deletions or rearrangements. Finally, we observed *TPST1* expression in patient lymphoblasts. Although mutations in regulatory elements, which may have more subtle effects on gene expression, remain a possibility, we conclude that *TPST1* is not involved in SDS.

In summary, we have refined the SDS locus to a 1.9 cM interval at 7q11, supported by observation of shared disease haplotypes across this interval. Construction of a physical map of the disease locus, based on a framework of ordered genetic markers, has facilitated genetic analysis, as well as mapping of candidate genes in the critical genetic interval. One of these, *TPST1*, was excluded as the causative gene for SDS by mutation analysis. In the future, integration of our map with the increasingly available genomic sequence of the region will allow a more comprehensive characterisation of candidate genes for SDS. Ultimately, this work should lead to identification of disease-causing mutations, providing for a more accurate diagnosis and better understanding of this complex disorder.

Acknowledgments

We are grateful for the co-operation of the patients with SDS, their families, and their physicians. We thank the Canadian Institutes of Health Research (CIHR) Genome Resource Facility at HSC for YAC and BAC library screening. We also acknowledge support from Shwachman-Diamond Canada, Shwachman Support of Great Britain, The Harrison Wright Appeal, Shwachman-Diamond Support of Australia, Shwachman-Diamond Support International, and the CIHR. JM Rommens is a member of the Canadian Genetic Diseases Network (CGDN). M Popovic received a joint CGDN-HSC Award and an Ontario Graduate Scholarship. S Goobie received a Postgraduate Scholarship from the Natural Sciences and Engineering Research Council of Canada.

References

- 1 Goobie S, Popovic M, Morrison J et al: Shwachman-Diamond syndrome with exocrine pancreatic dysfunction and bone marrow failure maps to the centromeric region of chromosome 7. *Am J Hum Genet* 2001; **68**: 1048–1054.
- 2 Ginzberg H, Shin J, Ellis L et al: Shwachman syndrome: phenotypic manifestations of sibling sets and isolated cases in a large patient cohort are similar. *J Pediatr* 1999; **135**: 81–88.
- 3 Ginzberg H, Shin J, Ellis L et al: Segregation analysis in Swachman-Diamond syndrome: evidence for recessive inheritance. *Am J Hum Genet* 2000; **66**: 1413–1416.
- 4 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-independent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.
- 5 Ouyang Y-B, Lane WS, Moore KL: Tyrosylprotein sulfotransferase: purification and molecular cloning of an enzyme that catalyzes tyrosine O-sulfation, a common posttranslational modification of eukaryotic proteins. *Proc Natl Acad Sci USA* 1998; **95**: 2896–2901.
- 6 Masuno M, Imaizumi K, Nishimura G et al: Shwachman syndrome associated with de novo reciprocal translocation t(6;12)(q16.2;q21.2). *J Med Genet* 1995; **32**: 894–895.
- 7 Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1215.
- 8 Rozen S, Skaletsky HJ: Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html
- 9 Vignal A, Gyapay G, Hazan J et al: Nonradioactive multiplex procedure for genotyping of microsatellite markers; in Adolph KW (ed): *Methods in molecular genetics*. vol 1, Academic Press, San Diego, 1993, pp 211–221.
- 10 Little S: Amplification-refractory mutation system (ARMS) analysis of point mutations; in Dracopoli NC, Haines JL, Korf BR et al. (eds): *Current protocols in human genetics*. vol 2, John Wiley & Sons, New York, 2000, pp 9.8.1–9.8.12.
- 11 Hudson TJ: Colony hybridization to screen for microsatellites; in Dracopoli NC, Haines JL, Korf BR et al. (eds): *Current protocols in human genetics*. vol 1, John Wiley & Sons, New York, 2000, pp 2.3.1–2.3.5.
- 12 Scherer SW, Tompkins BJ, Tsui LC: A human chromosome 7-specific genomic DNA library in yeast artificial chromosomes. *Mamm Genome* 1992; **3**: 179–181.
- 13 Osoegawa K, Woon PY, Zhao B et al: An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 1998; **52**: 1–8.
- 14 Liu Q, Thorland EC, Heit JA, Sommer SS: Overlapping PCR for bidirectional PCR amplification of specific alleles: a rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes. *Genome Res* 1997; **7**: 389–398.
- 15 MacDonald RJ, Swift GH, Przybyla AE, Chirgwin JM: Isolation of RNA using guanidinium salts. *Methods Enzymol* 1987; **152**: 219–227.
- 16 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.
- 17 Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. GenBank. *Nucleic Acids Res* 2000; **28**: 15–18.
- 18 Stewart EA, McKusick KB, Aggarwal A et al: An STS-based radiation hybrid map of the human genome. *Genome Res* 1997; **7**: 422–433.
- 19 de la Puente A, Velasco E, Perez Jurado LA et al: Analysis of the monomeric alphoid sequences in the pericentromeric region of human chromosome 7. *Cytogenet Cell Genet* 1998; **83**: 176–181.
- 20 Hudson TJ, Stein LD, Gerety SS et al: An STS-based map of the human genome. *Science* 1995; **270**: 1945–1954.
- 21 Bouffard GG, Idol JR, Braden VV et al: A physical map of human chromosome 7: an integrated YAC contig map with average STS spacing of 79 kb. *Genome Res* 1997; **7**: 673–692.
- 22 Deloukas P, Schuler GD, Gyapay G et al: A physical map of 30,000 human genes. *Science* 1998; **282**: 744–746.

- 23 Green ED, Idol JR, Mohr-Tidwell RM *et al*: Integration of physical, genetic and cytogenetic maps of human chromosome 7: isolation and analysis of yeast artificial chromosome clones for 117 mapped genetic markers. *Hum Mol Genet* 1994; **3**: 489–501.
- 24 Collins A, Frezal J, Teague J, Morton NE: A metric map of humans: 23,500 loci in 850 bands. *Proc Natl Acad Sci USA* 1996; **93**: 14771–14775.
- 25 Waye JS, England SB, Willard HF: Genomic organization of alpha satellite DNA on human chromosome 7: evidence for two distinct alphoid domains on a single chromosome. *Mol Cell Biol* 1987; **7**: 349–356.
- 26 Dunham I, Lengauer C, Cremer T, Featherstone T: Rapid generation of chromosome-specific alphoid DNA probes using the polymerase chain reaction. *Hum Genet* 1992; **88**: 457–462.
- 27 Vortkamp A, Thias U, Gessler M *et al*: A somatic cell hybrid panel and DNA probes for physical mapping of human chromosome 7p. *Genomics* 1991; **11**: 737–743.
- 28 Petrukhin K, Fischer SG, Pirastu M *et al*: Mapping, cloning and genetic characterization of the region containing the Wilson disease gene. *Nat Genet* 1993; **5**: 338–343.
- 29 Figus A, Angius A, Loudianos G *et al*: Molecular pathology and haplotype analysis of Wilson disease in Mediterranean populations. *Am J Hum Genet* 1995; **57**: 1318–1324.
- 30 Dib C, Faure S, Fizames C *et al*: A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 1996; **380**: 152–154.
- 31 Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 1998; **63**: 861–869.
- 32 Horvath JE, Bailey JA, Locke DP, Eichler EE: Lessons from the human genome: transitions between euchromatin and heterochromatin. *Hum Mol Genet* 2001; **10**: 2215–2223.
- 33 Eichler EE: Segmental duplications: what's missing, misassigned, and misassembled - and should we care? *Genome Res* 2001; **11**: 653–656.
- 34 Reich DE, Cargill M, Bolk S *et al*: Linkage disequilibrium in the human genome. *Nature* 2001; **411**: 199–204.
- 35 Yu A, Zhao C, Fan Y *et al*: Comparison of human genetic and sequence-based physical maps. *Nature* 2001; **409**: 951–953.
- 36 Zhang Q-H, Ye M, Wu X-Y *et al*: Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells. *Genome Res* 2000; **10**: 1546–1560.
- 37 Horii A, Han H-J, Sasaki S, Shimada M, Nakamura Y: Cloning, characterization and chromosomal assignment of the human genes homologous to yeast PMS1, a member of mismatch repair genes. *Biochem Biophys Res Commun* 1994; **204**: 1257–1264.
- 38 Li X-A, Kokame K, Okubo K *et al*: Cloning and characterization of a novel human gene encoding a zinc finger protein with 25 fingers. *Biochim Biophys Acta* 1999; **1489**: 405–412.
- 39 Olivier M, Aggarwal A, Allen J *et al*: A high-resolution radiation hybrid map of the human genome draft sequence. *Science* 2001; **291**: 1298–1302.
- 40 Tu L, Murphy PG, Li X, Tedder TF: L-selectin ligands expressed by human leukocytes are HECA-452 antibody-defined carbohydrate epitopes preferentially displayed by P-selectin glycoprotein ligand-1. *J Immunol* 1999; **163**: 5070–5078.
- 41 Kehoe JW, Bertozzi CR: Tyrosine sulfation: a modulator of extracellular protein-protein interactions. *Chem Biol* 2000; **7**: R57–R61.
- 42 Daniele A, Parenti G, d'Addio M, Andria G, Ballabio A, Meroni G: Biochemical characterization of arylsulfatase E and functional analysis of mutations found in patients with X-linked chondrodysplasia punctata. *Am J Hum Genet* 1998; **62**: 562–572.
- 43 Iida A, Sekine A, Saito S *et al*: Catalog of 320 single nucleotide polymorphisms (SNPs) in 20 quinone oxidoreductase and sulfotransferase genes. *J Hum Genet* 2001; **46**: 225–240.
- 44 Wang DG, Fan JB, Siao CJ *et al*: Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998; **280**: 1077–1082.
- 45 Kunz J, Scherer SW, Klawitz I *et al*: Regional localization of 725 human chromosome 7-specific yeast artificial chromosome clones. *Genomics* 1994; **22**: 439–448.