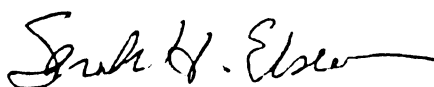This is to certify that the
dissertation entitled

MOLECULAR ANALYSIS OF SMITH-MAGENIS SYNDROME:
MAPPING, CANDIDATE GENE IDENTIFICATION, AND
PRELIMINARY CHARACTERIZATION OF RAI1

presented by

REBECCA E. SLAGER

has been accepted towards fulfillment
of the requirements for the

__Ph.D.__ degree in __Genetics__

Smith H. Elser
Major Professor's Signature

_3/22/04_
Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
AUG 2 4 2005		

6/01 c:/CIRC/DateDue.p65-p.15

MOLECULAR ANALYSIS OF SMITH-MAGENIS SYNDROME: MAPPING,
CANDIDATE GENE IDENTIFICATION, AND PRELIMINARY
CHARACTERIZATION OF *RAI1*

By

Rebecca E. Slager

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Genetics program

2004

ABSTRACT

MOLECULAR ANALYSIS OF SMITH-MAGENIS SYNDROME: MAPPING, CANDIDATE GENE IDENTIFICATION, AND PRELIMINARY CHARACTERIZATION OF *RAI1*

By

Rebecca E. Slager

Smith-Magenis syndrome (SMS) is a multiple congenital anomalies, mental retardation syndrome typically associated with a microdeletion of human chromosome 17 band p11.2. SMS patients display characteristic physical abnormalities as well as developmental delay, sleep disturbance, and self-injurious behaviors. The research focus in the Elsea laboratory is to understand the molecular basis of SMS and which gene or genes contribute to the phenotype of this complex physical and neurobehavioral disorder. Previous to this work, a great deal of effort was put forth to define the deletion interval and to begin to map genes, ESTs, and genomic markers to the SMS region of 17p. My research project focused on fine-mapping the smallest deleted region along 17p11.2 which could still reproduce the entire SMS phenotype, or what is termed the SMS critical interval. Through a combination of *in silico* analysis of high-throughput genome sequence and hybridization mapping of genes and expressed sequence tags (ESTs) to 17p11.2 genomic clones, a physical and transcription map of the SMS critical region was created. Mapping of genes and ESTs to the contig was reconciled with naturally-occurring patient deletions by Southern hybridization to digested DNA from a somatic cell hybrid panel of overlapping 17p deletions. The physical map of ~1.5 Mb SMS critical interval consisted of a minimum tiling path of 18 genomic clones and the transcription map contained 17 known genes, 12 ESTs, and six genomic markers. Recent

work by other members of the Elsea laboratory has re-defined the SMS critical interval as a ~950 kb region containing ~18-22 genes.

Several promising candidate genes were identified within the SMS critical interval and my research focused on the characterization of one particular gene, retinoic acid induced 1 (*RAI1*). This large, novel protein is highly expressed in the brain and shares some sequence homology with a transcriptional coactivator, though the cellular role of RAI1 is currently unknown. As we did not have the ability to assess the functional role of *RAI1*, we began a mutation screen of *RAI1* in several patients with an SMS phenotype but no detectable 17p11.2 deletion. We hypothesized that a mutation in a single gene in these patients may produce an SMS phenotype. Our studies subsequently revealed dominant, deleterious *RAI1* mutations in four patients. These findings strongly suggest that haploinsufficiency of the RAI1 protein is sufficient to produce the craniofacial and neurobehavioral features seen in SMS patients.

In order to assess Rai1 haploinsufficiency in a live animal model, a gene-targeting construct was developed to remove *Rai1* expression in mice and embryonic stem cells were subsequently identified which contained the correctly targeted construct. In addition, we assessed *Rai1* dosage sensitivity by constructing stable *Rai1* BAC transgenic lines and performed a preliminary physical assessment of these mice. Preliminary measurements of weight and total body length suggest that animals with higher *Rai1* copy numbers may weigh less and have a smaller body length than normal littermates.

I would like to dedicate this work to my grandfather, Dr. Robert E. Lucas, who received his Ph.D. from Michigan State University in 1947.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Chapter IV. *In vivo* evaluation of mouse *Rai1*

Chapter V. Discussion

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

aa:	amino acid
ABI:	Applied Biosystems
ADHD:	attention deficit hyperactivity disorder
AGS:	Alagille syndrome
ATPAF2:	ATP synthase mitochondrial F1 complex assembly factor 2
AS:	Angelman syndrome
ASP:	affected sib pair
BAC:	bacterial artificial chromosome
BAZ1B:	bromodomain adjacent to zinc finger domain, 1B
bp:	base pair
COPS3:	COP9 homolog subunit 3
CREBBP:	Creb-binding protein
Cyln2:	cytoplasmic linker 2
DGS:	DiGeorge syndrome
DHPLC:	denaturing high-performance liquid chromatography
DRG2:	developmentally-regulated GTP-binding protein 2
ELN:	Elastin
ES:	embryonic stem
Fgf8:	Fibroblast growth factor 8
FISH:	fluorescent *in situ* hybridization
FLII:	flightless I homolog
GFP:	green fluorescent protein
GTSF:	Genomics Technology and Support Facility
HTGS:	high-throughput genome sequence
IHC:	immunohistochemistry
ISH:	*in situ* hybridization
JAG1:	Jagged1
kb:	kilobase
LCR:	low-copy repeat
LIMK1:	Lim-kinase 1
LLGL1:	lethal giant larvae homolog 1
LOD:	logarithm of odds
M-RIP:	myosin phosphatase-Rho interacting protein
MYO15A:	myosin XVA
NAHR:	non-allelic homologous recombination
ng:	nanograms
NT5M:	5',3'-nucleotidase, mitochondrial
NTF:	non-transgenic female
NTM:	non-transgenic male
PAC:	P1-artificial chromosome
PBS:	phosphate-buffered saline
PCR:	polymerase chain reaction
PEMT2:	phosphatidylethanolamine N-methyltransferase

pg:	picograms
PHD:	plant homeodomain
PWS:	Prader-Willi syndrome
RA:	retinoic acid
RAI1:	retinoic acid induced 1
RARE:	retinoic acid response element
RASD1:	RAS, dexamethasone-induced 1
REM:	rapid eye movement
SCA2:	spinocerebellar ataxia type 2
SHMT1:	serine hydroxymethyl transferase 1
SMS:	Smith-Magenis syndrome
SNP:	single nucleotide polymorphism
SREBF1:	sterol regulatory element binding transcription factor 1
SVAS:	supravalvular aortic stenosis
Tbx1:	T-box containing transcription factor 1
TCF20:	Transcription factor 20
TGCE:	temperature gradient capillary electrophoresis
TF:	transgenic female
TM:	transgenic male
TOM1L2:	target of myb1-like 2
TOP3A:	topoisomerase III alpha
U of M TAMC:	University of Michigan Transgenic Animal Core
UBE3A:	E6-AP ubiquitin protein ligase
UTR:	untranslated region
VCFS:	velo-cardio-facial syndrome
WS:	Williams syndrome
WSCR:	Williams syndrome critical interval
YAC:	yeast artificial chromosome

Chapter I.

Introduction

Microdeletion syndromes and genomic disorders

Smith-Magenis syndrome (SMS) is a multiple congenital anomalies, mental retardation syndrome typically associated with a microdeletion of human chromosome 17 band p11.2 (Smith et al. 1982; Smith et al. 1986; Greenberg et al. 1991; Greenberg et al. 1996). Microdeletion syndromes result from small deletions of chromosomal material and form a subset of chromosomal abnormalities which traditionally account for 6%-10% of all infants born with severe congenital anomalies (Shapira 1998). Diagnosis of a particular microdeletion syndrome is typically made by molecular cytogenetic procedures such as fluorescent *in situ* hybridization (FISH), as the deletions can be undetectable even by means such as high-resolution chromosomal banding analysis (prometaphase or metaphase spreads to 550 band resolution). FISH analysis relies on the hybridization of a fluorescently labeled probe that is specific for a region of a chromosome; the absence of one chromosomal copy of the signal in a metaphase spread is diagnostic for a deletion. FISH can also be useful in the confirmation of a clinical diagnosis and the detection of patients with different deletion sizes (Shapira 1998). As a microdeletion generally occurs *de novo* on only one member of a homologous pair of chromosomes, the mode of inheritance of the microdeletion syndrome is considered to be autosomal dominant.

As molecular cytogenetics improves and more complex chromosomal rearrangements can be detected, a new discipline has emerged to understand the

underlying mechanism of deletions, duplications, translocations, and inversions, which fall under the broad term of genomic disorders. Through the completion of the human genome project, new sequence data indicate that many of the chromosomal regions throughout the genome that are prone to rearrangement are flanked by chromosome-specific low-copy repeats (LCRs). The structure of these LCR segments can be simple or complex and may encompass various genes, gene fragments, pseudogenes, and retroviral sequences (Ji et al. 2000; Stankiewicz et al. 2003). Many of these LCRs are present in >2-10 copies and tend to be preferentially located in centromeric and telomeric regions, and the genomic distance between the LCR blocks is typically kilobases to megabases (Ji, Eichler et al. 2000; Stankiewicz et al. 2002). LCR repeat blocks of ~10-400 kb are highly similar (>95%) and have the ability to act as substrates for non-allelic homologous recombination (NAHR) or unequal crossing over during meiosis, which results from the mis-pairing of the repeated fragments (Stankiewicz, Shaw et al. 2003). NAHR can occur between homologous chromosomes, sister chromatids, or within a single chromatid to produce unequal products of recombination or a "looping out" of genetic material (Stankiewicz and Lupski 2002). Several genetic disorders, including microdeletion and duplication syndromes or other disorders caused by large chromosomal rearrangements, result from NAHR sponsored by the mispairing of flanking LCRs. The origin of many of these LCR segments is thought to be duplication events which occurred during primate speciation approximately 35-50 million years ago (Stankiewicz and Lupski 2002), though how these sequences became fixed within the human genome is currently unknown.

The complex nature of chromosome-specific LCRs has recently been under intense investigation following the availability of human draft and finished genomic sequence. Many genomic regions, including the pericentromic regions of chromosome 17, did not have correct sequence alignment or orientation due to the high similarity between the LCRs, which caused the misassembly of several bacterial artificial chromosome (BAC) and P1 artificial chromosome (PAC) clones. This was especially true of repeated LCR blocks of ~200 kb, which is the size of many sequenced BACs (Stankiewicz, Shaw et al. 2003). Many reviews now exist which analyze the structure of the LCRs flanking known microdeletions as well as the possible mechanism of NAHR (Ji, Eichler et al. 2000; Park et al. 2002; Stankiewicz and Lupski 2002; Stankiewicz, Shaw et al. 2003). As our lab is interested primarily in the molecular basis of SMS, this work will highlight the LCR repeat regions which flank the SMS common deletion breakpoints within 17p11.2, termed the proximal and distal SMS-REPs, or SMS-REPP and SMS-REPD, as well as a third middle REP (SMS-REPM). The SMS-REPs contain at least 14 genes and pseudogenes including the keratin *(KER)* gene cluster, *CLP*, and *TRE*. Many of these genes also map to multiple locations throughout the genome, including the q arm of chromosome 17 (Park, Stankiewicz et al. 2002). The SMS-REPs are >98% similar repeat blocks of ~200kb which are believed to have arisen from a progenitor SMS-REP with nearly the same structure as the SMS-REPP, nearly ~40-65 million years old during the divergence of Old and New World monkeys (Park, Stankiewicz et al. 2002; Stankiewicz and Lupski 2002). While the SMS-REPP and SMS-REPD occur in the same orientation, various complex rearrangements, including two terminal deletions, an interstitial deletion, and an inversion which occurred throughout

the evolution of the SMS-REPs have been proposed to explain the existence of the SMS-REPM, which occurs in the opposite orientation (Park, Stankiewicz et al. 2002; Stankiewicz and Lupski 2002). Interestingly, sequence homology comparisons between mouse and human genomes show that the SMS-REPs are present at breaks in the synteny of these genomes (Stankiewicz and Lupski 2002). The highly similar restriction pattern of several sequenced BACs from the human genome project which comprise the SMS-REPs are shown in the Southern hybridization in Figure 1, panel (a). As shown in the schematic in Figure 1, panel (b), *de novo* deletions and duplications within 17p11.2 are likely due to interchromosomal NAHR events involving the SMS-REPs (Chen et al. 1997). The SMS-REPP and the SMS-REPD, which flank the ~4 Mb SMS common deletions are in the same orientation and unequal crossing over of these repeat blocks can sponsor meiotic NAHR which result in the SMS common deletion (Chen, Manian et al. 1997). Less frequently, the inverted SMS-REPM can also act as a substrate for NAHR, as at least five reported deletions have been analyzed which involve the SMS-REPM (Park, Stankiewicz et al. 2002; Stankiewicz, Shaw et al. 2003). Several patients with the SMS common deletion were identified by the presence of a unique junction fragment, the putative remnant from an unequal crossing over event (Chen, Manian et al. 1997). In-depth molecular analysis of the recombinant junctions from SMS patients with the common deletion demonstrate that ~50% of NAHR events occurred within the *KER* gene cluster (Bi et al. 2003). Patients with the reciprocal duplication of 17p11.2 have also recently been identified and characterized (Potocki et al. 2000). As SMS patients usually display a number of physical abnormalities and self-abusive behaviors (Greenberg, Guzzetta et al. 1991; Smith et al. 1998), the phenotype of these dup(17)(p11.2 p11.2)

4

Figure 1. Repetitive elements involved in the SMS common 17p11.2 deletion and dup(17)(p11.2p11.2) duplication.

(a) The similar restriction pattern of the SMS-REPs is represented by Southern hybridization. DNA from BACs mapping to the proximal REP (bc676C11 and bc434D2, shown in pink), distal REP (bc198H15, pc48J14 and bc219A15, shown in purple) and middle REP (bc158M20, bc581M24, pc37N07 and bc34O10, shown in black), as well as a BAC containing unique sequence (bc815I9, shown in blue) was digested with EcoRI, electrophoresed on a 1% TAE gel, and transferred to a nylon membrane. The Southern blot was probed with a >10 kb fragment from bc198H15, which hybridized to several bands in BACs from all three of the SMS-REP sequences and highlights the extremely similar restriction pattern of these BACs. As a negative control, bc198H15 BAC fragment did not bind to the unique genomic sequence in bc815I9. In-depth sequence analysis reveals that the SMS-REPS are >98% identical (Stankiewicz et al. 2002).

(b) Meiotic non-allelic homologous recombination on human chromosome 17p11.2 is represented (adapted from Potocki et al. 2000), which results in two products: the SMS common deletion as well as the reciprocal duplication of the 17p11.2 region. The SMS-REPP (pink) and SMS-REPD (purple) flank unique genomic sequence which includes the FLII and ZNF179 genes. Homologous recombination can occur between the ~95% identical SMS-REPP and the SMS-REPD, which produces the SMS common deletion and dup(17)(p11.2p11.2), both of which can be identified by unique junction fragments. The SMS-REPM, which occurs in the opposite orientation as the SMS-REPP and the SMS-REPD and may sponsor unusual 17p11.2 deletions, is not represented in this diagram.

Images in this dissertation are presented in color.

bc676C11
bc434D2
bc198H15
bc81519
pc48J14
bc158M20
bc581M24
pc37N07
bc34O10
bc219A15

b.

SMS-REPP

ZNF179

FLII

SMS-REPD

PMP22

cen

A
B

C
D

tel

A'
B'
C'
D'

A
B

C

B'

C'
D'

Junction
fragment

dup(17)(p11.2p11.2)

A'
D

Junction
fragment

SMS deletion
del(17)(p11.2p11.2)
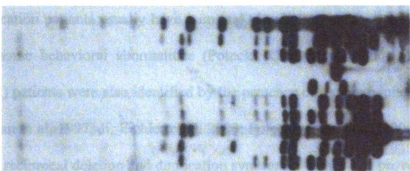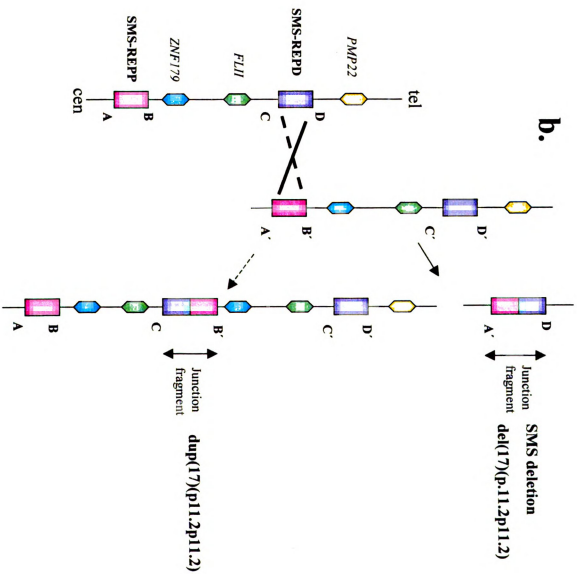
patients is generally considered to be less severe (Potocki, Chen et al. 2000). The duplication patients usually have a normal physical appearance, mild mental retardation and some behavioral abormalities (Potocki, Chen et al. 2000). These dup(17)(p11.2 p11.2) patients were also identified by the presence of a unique junction fragment (Chen, Manian et al. 1997; Ji, Eichler et al. 2000; Potocki, Chen et al. 2000). The existence of these reciprocal deletion and duplication syndromes within 17p provides evidence for the model of NAHR sponsored by complex genome architecture (Chen, Manian et al. 1997; Potocki, Chen et al. 2000) and also provides a unique opportunity to identify particularly dosage-sensitive genes within the 17p11.2 region.

SMS clinical summary

Smith-Magenis syndrome (SMS) is a multiple congenital anomalies/mental retardation syndrome associated with an interstitial deletion on chromosome 17 involving band p11.2 (Smith, L. et al. 1982; Smith, McGavran et al. 1986; Stratton et al. 1986; Greenberg, Guzzetta et al. 1991). The birth prevalence of SMS is estimated to be approximately 1:25,000 (Greenberg, Guzzetta et al. 1991), although SMS is likely underdiagnosed due to the fact that it is a recently-described syndrome and its phenotype can be subtle. Many cases have been identified in the last 10 years as a result of high-resolution cytogenetic techniques such as FISH. SMS is known to affect individuals of all ethnic and racial backgrounds. A detailed phenotypic summary of SMS is available on the world wide web at the GeneReviews website (www.genetests.org), through executing a search on this website for Smith-Magenis syndrome. This website is overseen by a board of professionals who update the site regularly and provide accurate,

referenced information about the clinical and molecular aspects of SMS. Most of the SMS clinical information below has been adapted from the SMS entry at GeneReviews:

In infancy, there are very subtle signs that a baby may have SMS. Usually, the infant is born at term and presents with a normal birth height and weight. As a baby, an SMS child may display dysmorphic facies, characterized by midface hypoplasia, a short, up-turned nose, and a "tented" upper lip, although these physical features may be so subtle as to be overlooked. Infants with SMS are often described as "perfect babies" who sleep for long periods of time and cry infrequently. Some feeding difficulties are common in SMS infants, as well as poor suck, and gastroesophogeal reflux. Nearly 100% of SMS infants are hypotonic and lethargic (Greenberg, Lewis et al. 1996).

The physical features of an individual with SMS become more pronounced in early childhood. In addition to the typical SMS facial appearance noted above, children with SMS may have ocular abnormalities, including myopia, strabismus, or microcornea. In general, SMS children also have short stature as well as short hands and feet and some SMS children have mild scoliosis. Other common features of children with SMS include otolaryngologic difficulties, otitis media (often severe enough to induce ear tube placement), hearing loss, and hypercholesterolemia (Smith, McGavran et al. 1986; Smith et al. 1990; Greenberg, Guzzetta et al. 1991). In general, oral sensory motor dysfunction is a major concern, especially as many SMS children have a marked hoarse voice and significant speech delay. However, with the proper intervention, an SMS child can

develop verbal speech by school age. Cardiac, renal anomalies and cleft palate have been noted to occur in smaller percentage of patients (www.genetests.org).

The sleep disturbances in an SMS child shift from prolonged napping in infancy to a repeated pattern of naps and awakenings (De Leersnyder et al. 2001). As an SMS child ages, the number of naps increases and sleep at night decreases. In general, there is an overall decrease in rapid eye movement (REM) sleep. Recently, SMS children have also been documented to have an inversion of the normal circadian rhythm of melatonin, which is the most likely cause of their sleep disturbances (Potocki et al. 2000; De Leersnyder, De Blois et al. 2001). Little is known about how the melatonin secretion and distribution pathway is affected in these individuals, though a study of SMS children given a specific β_1–adrenergic agonist (acetbutolol) as well as an evening dose of melatonin demonstrated markedly increased hours of sleep and decreased awakenings (De Leersnyder et al. 2001). Some overall improvement in the inappropriate behaviors of these children was also noted (De Leersnyder, de Blois et al. 2001).

SMS children are often diagnosed with some form of developmental delay and mild to moderate mental retardation. During school age, the behavioral aspects of the SMS phenotype, which include attention deficit and hyperactivity, attention-seeking behavior, tantrums, aggression and self-injury, become pronounced and may escalate through puberty (Smith, Dykens et al. 1998). Several self-injurious behaviors such as self-biting, skin picking, and self-hitting are common in SMS patients. Two behaviors unique to most SMS patients are termed onychotillomania (pulling out of nails) and

polyemboilokomania (insertion of objects into bodily orifices) (Smith, Dykens et al. 1998; Finucane et al. 2001). Several stereotypic behaviors such as self-hugging, teeth grinding, and body rocking are also common to many SMS patients (Finucane, Dirrigl et al. 2001).

Following puberty, the physical aspects of SMS, especially the facial features, such as midface hypoplasia, prognathism, synophrys, and heavy brows tend to become more pronounced. Scoliosis may also become more severe with age. The neurobehavioral phenotype may worsen during adolescence, though the aggressive and self-injurious behaviors eventually may lessen through adulthood (Smith, Dykens et al. 1998). An ongoing life history study of SMS patients conducted at the National Institutes of Health (NIH) and spearheaded by Ann Smith should provide valuable information about adult individuals with SMS.

Determination of the molecular basis of microdeletion syndromes

Patients diagnosed with a microdeletion syndrome present with a characteristic phenotype, which typically results from reduction of functional proteins consequent to the loss of one copy of a gene or genes that is/are contained within the deleted chromosomal region. Often this is due to haploinsufficiency for a dosage-sensitive gene product or products, though it is also possible that hemizygosity for a recessive gene may produce a phenotype if the remaining copy is mutated or disrupted in some way. Several microdeletion syndromes exist whose primary phenotype arises from the disruption of a single gene, such as Alagille syndrome (AGS), which is caused by dominant mutations or

deletion of one copy of the human Jagged1 (*JAG1*) gene (Li et al. 1997) and Rubenstein-Taybi syndrome, which is caused by haploinsufficiency of the Creb-binding protein (*CREBBP*) (Petrij et al. 1995; Oda et al. 1997).

For several other microdeletion syndromes, it is possible that haploinsufficiency of a single gene is not enough to produce the entire phenotype (Shapira 1998). It is likely that these are contiguous gene syndromes. This term, introduced by Schmickel in 1986, is used to describe the involvement of multiple functionally-unrelated genes that lie close to one another on a chromosome (Schmickel 1986) and the complex phenotype that arises from gene dosage imbalances of several genes associated with chromosomal deletions or duplications. A true contiguous gene syndrome contains several genes that may contribute to the phenotype and at least one or more genes that follows a Mendelian pattern of inheritance which can be traced in the general population (Schmickel 1986). An example of a contiguous gene syndrome is Williams syndrome (WS), which is associated with a common deletion of 7q11.2 that encompasses the elastin (*ELN*) locus (Francke 1999). The ~50 kb *ELN* maps to the center region of Williams syndrome critical region (WSCR) and encodes the protein tropoelastin. Haploinsufficiency of *ELN* alone is sufficient to account for the cardiac and vascular abnormalities that are observed in WS patients but not for other features associated with the syndrome such as hypercalcemia, dysmorphic facial features, and a distinct cognitive profile. Patients who only harbor small deletions or point mutations within the *ELN* gene present with one of the two autosomal dominant cardiac disorders (typically supravavular aortic stenosis or cutis laxa) but not the other physical and neurobehavioral characteristic aspects of WS

(Francke 1999; Peoples et al. 2000). The role of other genes within the WSCR are currently under investigation in order to understand their contribution to the full WS phenotype. At this time, it is not yet known whether SMS [del(17)p11.2] is a contiguous gene syndrome, though data discussed in this work suggest that the majority of the SMS phenotype may be due to haploinsuffiency of one gene, retinoic acid induced 1 or *RAI1*.

Determining the primary gene or genes involved in producing the phenotype of a particular microdeletion syndrome is an extremely complex process and multiple approaches are combined to identify and analyze promising candidate genes. Several labs have used techniques similar to those of our own group to identify the genes involved in AGS and WS. Increasingly (as discussed below), animal models are becoming an invaluable resources in understanding the *in vivo* effects of gene dosage and how haploinsufficency of a single gene can have pleiotropic effects. This work will discuss AGS and WS as particular models of microdeletion syndrome analysis and techniques we have applied to understand the molecular basis of SMS. AGS has aspects of a microdeletion syndrome as well as a single gene disorder. The study of WS is also a useful model for SMS, as genes other than *ELN* that map to the WSCR are currently under investigation in order to understand this contiguous gene disorder. Our lab is using similar approaches to understand how *RAI1* haploinsufficiency can produce the multiple phenotypic effects of SMS and to decipher the role of other hemizygous genes in the deleted region.

AGS (OMIM #118450) is an autosomal dominant disorder now commonly diagnosed using at least three of the five major clinical characteristics first described by Alagille in 1975: cardiac disease including pulmonic valvular stenosis and peripheral pulmonary arterial stenosis, cholestasis, skeletal abnormalities including "butterfly" vertebrae, ocular abnormalities, and a characteristic facies consisting of a broad forehead, pointed mandible, deep-set eyes, and an unusual bulbous tip on the nose (Alagille et al. 1975; Krantz et al. 1997; Oda, Elkahloun et al. 1997). Many AGS patients present with neonatal jaundice and cholestasis due to an insufficient number of bile ducts, which are identified on liver biopsy (Gridley 2003). While some degree of mental retardation has been reported in a subset of patients, most individuals with AGS have normal intelligence (Oda, Elkahloun et al. 1997). Many familial cases of AGS exist, both those with normal karyotypes and those who harbor interstitial deletions of 20p12. The latter finding strongly suggested that the locus for AGS mapped within this deletion region (Oda, Elkahloun et al. 1997). Further linkage analysis in an AGS family with no detectable deletion confirmed the locus mapped to 20p, between the genetic markers D20S59 and D20S65 (Oda, Elkahloun et al. 1997). In depth cytogenetic analysis also revealed that two AGS families were carrying a balanced translocation with t(2;20)(q21.3;p12) which helped to define an preliminary centromeric boundary of the AGS critical interval. The telomeric boundary was defined by FISH analysis with a clone for the *SNAP* gene (Oda, Elkahloun et al. 1997). Thus, the critical interval for AGS was defined as an ~1.3 Mb interval between *SNAP* and D20S186 (Oda, Elkahloun et al. 1997). A preliminary 3.7 Mb genomic contig containing this chromosomal region was then created using yeast artificial chromosomes (YACs) (Pollet et al. 1995). Using information generated from

this contig as well as high-resolution cytogenetic analysis of two AGS patients with submicroscopic 20p12 deletions, one group reduced the AGS critical region to ~250 kb and a new overlapping, genomic contig was created using bacterial artificial chromosomes or BAC clones (Oda, Elkahloun et al. 1997). This same group had also mapped a cDNA representing the human homolog of the rat *Jagged1* gene (*JAG1*) to this AGS critical region (Oda et al. 1997). *JAG1* is widely expressed and was an attractive AGS candidate gene because it encodes a ligand for the Notch transmembrane receptor, which has been reported to be critical for the determination of certain developmental cell fates (Oda, Elkahloun et al. 1997). The genomic structure of human *JAG1* was determined and detrimental dominant mutations (including frameshifts, and splice donor mutations) were identified within the coding region of *JAG1* in several AGS patients with no cytogenetic deletion (Oda, Elkahloun et al. 1997). A separate group also simultaneously identified distinct *JAG1* coding mutations in four unrelated patients (Li, Krantz et al. 1997). Once these mutations were reported, analysis of the genotype/phenotype profile of AGS intensified as well as an attempt to further understand the expression pattern and cellular role of the JAG1 protein. A recent extensive survey of the *JAG1* mutations in AGS summarized the types and frequencies of sequence changes found in patients. Briefly, the findings revealed: 72% of the mutations led to a premature stop codon, 15% were splice site mutations, 13% were missense mutations and 3-7% of patients carried deletions of the entire *JAG1* gene (Spinner et al. 2001). While the JAG1 protein does contain several well-conserved domains, mutations were found to occur throughout the gene and not within a particular domain (Piccoli et al. 2001). It is also important to note that *JAG1* mutations have only

been found in ~70% of AGS patients (Piccoli and Spinner 2001; Gridley 2003). The etiology of AGS in these patients is currently unknown, though it is possible that they may harbor cryptic mutations that may affect the regulation or expression of *JAG1*. However, extensive *JAG1* mutation screening has revealed no clear genotype/phenotype correlation. Analysis of AGS is also complicated by the high penetrance but variable expressivity of individuals carrying *JAG1* mutations (Gridley 2003). Familial mutations are common but phenotypic features may vary greatly between patients, and not all family members harboring presumably deleterious mutations may meet the clinical criteria of AGS (Gridley 2003). Another complicating factor in *JAG1* mutation screening is the clinical significance of AGS missense mutations. These standard criteria are typically used to determine whether a missense mutation is disease causing: a.) determination that the sequence change occurs within an evolutionary conserved amino acid, b.) segregation within a family with disease phenotype, and c.) absence of the change within the general population (Piccoli and Spinner 2001). Several studies have discovered particular *JAG1* missense mutations that segregate only with particular cardiac features, notably pulmonic stenosis or tetralogy of Fallot and hypoplatic pulmonary arteries but not the entire AGS syndromic features (Krantz et al. 1999). Another family harboring a particular *JAG1* missense mutation was found to exhibit hearing loss, vestibular defects, and congenital heart defects (Le Caignec et al. 2002). These important mutation studies demonstrate that there is much to be learned about the cellular role of JAG1 and that genetic modifiers may greatly affect the clinical phenotype of AGS. This hypothesis is supported by animal models of AGS (discussed below) (Gridley 2003).

The molecular analysis of WS (OMIM #194050) has many similarities to AGS as well as SMS. The main difference between WS and AGS is that haploinsufficiency of a single gene, *JAG1*, is thought to be responsible for the majority of the AGS phenotype, and multiple genes are believed to contribute to the complete WS phenotype. WS is associated with ~1.6 Mb heterozygous deletion of human chromosome 7 band q11.23 and diagnosis for WS is typically made using FISH with cosmid probes for *ELN* (Francke 1999). Physical features of WS include a flat nasal bridge, a short, up-turned nose, a long philtrum, full lips and lower cheeks and a characteristic small chin as well as a cardiac abnormalities (most frequently supravavular aortic stenosis or SVAS), hypercalcemia, constipation, and impaired vision (Osborne 1999). Most WS patients also have mild mental retardation, but the distinct behavioral and cognitive profile is perhaps the most intriguing aspect of WS, as it is composed of friendliness and anxiety as well as relatively high performance in linguistics but severely impaired visual-spatial abilities (Francke 1999; Osborne 1999). WS patients have motor delay and ~70 % suffer from attention deficit and hyperactivity disorder (ADHD) (Francke 1999; Osborne 1999). In order to understand the molecular mechanism of WS, as with SMS, deletions in WS patients were analyzed to determine the extent of a particular deletion and eventually to create a physical and transcription map of the deleted region (Francke 1999; Peoples, Franke et al. 2000). While unusual deletions identified in several SMS patients or those patients harboring overlapping deletions within 17p11.2 greatly enhanced the genomic and phenotypic analysis of the SMS critical interval (Bi et al. 2002; Vlangos et al. 2002), deletion sizes in WS were found to be extremely uniform (Perez Jurado et al. 1996;

Francke 1999). Similar to the SMS-REPs flanking the common deletion on 17p, LCRs of ~320 kb containing several genes, portions of genes, and pseudogenes, including *NCF1* and *GTF2I*, as well as members of the *PMS2* mismatch repair family of genes, also flank the WS deleted region (DeSilva et al. 1999; Francke 1999; Stankiewicz and Lupski 2002). While these repeated segments made assembly of BACs and PACs within the flanking region of the WS deletion difficult, the creation of fluid, overlapping physical and transcription maps of the unique sequence within the WSCR was relatively straightforward, using a combination of the molecular biological techniques and *in silico* data from the human genome project (Hockenhull et al. 1999; Peoples, Franke et al. 2000). Once the physical and transcription map of the WSCR had been created and ~19 transcripts were mapped to this region (Karmiloff-Smith et al. 2003), the next major challenge became to understand the genotype/phenotype correlation between haploinsufficiency for genes within the WSCR other than *ELN*. The phenotypic characteristics associated with deficiency of tropoelastin are cardiac abnormalities, hernias and possible premature aging (Karmiloff-Smith, Grant et al. 2003). The role of the other ~18 genes within the WSCR is still unknown, though Lim-kinase 1 (*LIMK1*) has been discussed as a possible candidate gene for the cognitive defects in WS, as this gene is highly expressed in neurons and has been shown to phosphorylate cofilin, which may be involved in regulation of actin reorganization (Karmiloff-Smith, Grant et al. 2003). However, several high-functioning SVAS patients with an above-average cognitive profile and no spatial impairment have been reported to be deleted for a portion of the WSCR, including *LIMK1*, so perhaps haploinsufficiency of *LIMK1* alone is not enough to produce the mental defects seen in WS patients (Hockenhull, Carette et al.

1999; Karmiloff-Smith, Grant et al. 2003). While to date, no genes other than *ELN* have been definitely implicated in the WS phenotype, it is interesting that a novel, putative transcriptional regulator, WBSCR9, now named *BAZ1B,* or WSTF, has been mapped to this critical region (Peoples et al. 1998). This novel gene contains numerous structural motifs, including a bromodomain, putative nuclear localization signals, nuclear receptor binding motifs, and a plant homeodomain or PHD finger domain (Peoples, Cisco et al. 1998; Franke et al. 1999), which is also predicted to be present in *RAI1*, the gene we believe to be responsible for many of the phenotypic characteristics of SMS. Also similar to *RAI1, BAZ1B* is widely expressed throughout development in human and mice (Peoples, Cisco et al. 1998). The PHD domain has been implicated in protein-protein interactions and is present in some proteins involved in chromatin-remodeling (Aasland et al. 1995). The structure of the PHD domain within the BAZ1B protein has been crystallized, revealing a core Cys_4-His-Cys_3 zinc binding domain with the ability to coordinate two Zn^{2+} (Pascual et al. 2000). Outside of the core zinc finger are two variable loops which are thought to confer specificity to binding ligands (Pascual, Martinez-Yamout et al. 2000). While highly speculative, it would be fascinating to discover a similar cellular function by *BAZ1B* and *RAI1* and perhaps a similar role in mental or behavioral development. In both WS and SMS, current assessment of promising candidate genes is being carried out using mouse models in order to link complex phenotypes to specific developmental pathways.

To further complicate the phenotypic and molecular analysis of a certain microdeletion syndrome, genomic imprinting may contribute to the overall phenotype.

Genomic imprinting is a normal process wherein a particular gene is "functionally hemizygous and retain[s] molecular memory of [its] parental origin throughout embyrogenesis using parental and allele-specific DNA methylation" (Khan et al. 1999). A great deal has been discerned about the normal imprinting process from classic studies of imprinting disorders such as the Angelman syndrome (AS) and Prader-Willi syndrome (PWS). Both of these disorders can be associated with a similar ~4 Mb deletions on chromosome 15q11-q13, although the former is typically derived from a maternal deletion and the latter by a paternal deletion (Mann et al. 1999; Ohta et al. 1999; Tsai et al. 1999; Butler 2002). Despite a common microdeletion, the phenotype and molecular basis of AS and PWS appear to be vastly different. AS patients display severe cognitive impairment, ataxia, and inappropriate laughter (Khan and Wood 1999), which most likely results from the lack of expression from the maternal allele of E6-AP ubiquitin protein ligase (*UBE3A*) (Kishino et al. 1997; Matsuura et al. 1997). *UBE3A* has been found to be paternally imprinted only in the brain (Rougeulle et al. 1997; Vu et al. 1997). Deleterious frameshift mutations in *UBE3A* identified in AS patients without a deletion indicate that this is a single-gene disorder (Kishino, Lalande et al. 1997; Matsuura, Sutcliffe et al. 1997), whereas PWS may be a contiguous gene disorder. The phenotype of PWS includes hyperphagia, severe obesity, hypotonia, short hands and feet, hypogonadism and a typical facies, and no patients have been identified to date with a single gene mutation (Vogels et al. 2002). Currently, several paternally expressed genes in the 15q11-q13 region are under investigation as PWS candidate genes (Cassidy et al. 2000). As SMS patients have been identified with both maternal and paternal deletions and the phenotypic characteristics of all patients are similar, imprinted genes are not believed to

contribute to the SMS phenotype (Greenberg, Guzzetta et al. 1991) and will not be discussed further in this work.

Animal models of microdeletion syndromes

The monoallelic microdeletion of chromosome 22q11.2 or del(22)q11 is thought to be one of the most common genomic disorder(s), estimated to occur in ~1:4000 live births (Lindsay et al. 1998; Yamagishi 2002). Several multiple congenital anomalies syndromes have been separately mapped to this region, including the DiGeorge syndrome (DGS), velo-cardio-facial syndrome (VCFS), as well as conotrucal anomaly face syndrome. Many of these disorders share overlapping clinical features. As molecular cytogenetics have improved, it was determined that many of this spectrum of disorders shared a common deletion region within 22q11.2 and is now known as 22q11 deletion syndrome (hereafter referred to as 22q11DS) (Yamagishi et al. 2003). While the clinical findings of 22q11DS are variable, approximately 75% of patients have congenital heart defects, typically of the cardiac outflow tract and aortic arch. Patients also have characteristic facies, immunodeficiency, hypocalcemia, and developmental and behavioral problems (Yamagishi and Srivastava 2003). To date, ~30 genes have been mapped to the 22q11DS critical interval and direct sequencing of candidate genes in DGS and VCFS patients with no detectable deletion has not revealed deleterious mutations, suggesting perhaps that haploinsufficiency of multiple genes may produce the phenotypic features (Yamagishi and Srivastava 2003). As 22q11DS is an extremely complex genomic disorder, animal models have played a crucial role in identifying the *in vivo* role of promising candidate genes. Several candidate genes, including *HIRA*, *UFD1L* and

CRKL, were separately targeted in mice and heterozygotes were not found to be haploinsufficient, though it remains a possibility that these genes may modify the human phenotype (Lamour et al. 1995; Pizzuti et al. 1997; Roberts et al. 1997; Wilming et al. 1997; Lindsay and Baldini 1998; Farrell et al. 1999; Yamagishi et al. 1999; Guris et al. 2001; Lindsay et al. 2001; Yamagishi et al. 2003). In order to reproduce the 22q11DS phenotype in an animal model, several groups have applied sophisticated genetic manipulations to remove the orthologous chromosomal region in mice (Vitelli et al. 2002). Briefly, this embryonic stem (ES) cell technology involves Cre-mediated *lox*P recombination (Vitelli, Lindsay et al. 2002). For example, in one targeting experiment, a *lox*P site was engineered into the *Hira* gene and one was inserted between the *GbpIbb* and the *Cld5* gene, ~150 kb centromeric to the adjacent *lox*P site (Vitelli, Lindsay et al. 2002). Cre recombinase specifically targets the genomic region between the loxP sites for deletion (Vitelli, Lindsay et al. 2002). Phenotypically, the mice carrying the heterozygous deletion of the chromosomal region orthologous to 22q11DS displayed somewhat more subtle features than human 22q11DS patients, though several significant phenotypic features were reproduced, such as aortic arch abnormalities, and parathyroid and neurobehavioral defects (Lindsay 2001; Paylor et al. 2001; Taddei et al. 2001). This same mouse ES cell engineering technology is currently being used in SMS studies by the Lupski research group at the Baylor College of Medicine to generate mice hemizygous for the region of mouse chromosome 11 which is orthologous to the SMS common deletion region (Walz et al. 2002; Walz et al. 2003), as well as to create nested deletions within this genomic region, in order to more precisely map phenotypic effects of certain genes.

In parallel with large-scale analysis of the orthologous chromosomal region to the 22q11DS critical region in mice, targeted disruptions of single candidate genes can provide valuable information about the phenotypic contribution of each particular gene. Through gene-targeting, disruption of one promising candidate gene, the T-box containing transcription factor (*Tbx1*), was also found to produce aortic arch defects (Lindsay, Vitelli et al. 2001). *Tbx1* is a member of a family of proteins containing a T-box, a domain which has been implicated in DNA-binding as well as protein-protein interactions (Lindsay, Vitelli et al. 2001; Yamagishi and Srivastava 2003). Mice hemizygous for *Tbx1* had aortic arch defects and homozygous mice displayed an even more severe phenotype, reproducing most of the cardiac abnormalities seen in 22q11DS as well as craniofacial abnormalities, cleft palate, and hyploplasia of the thymus and parathyroid glands (Lindsay and Baldini 1998; Guris, Fantes et al. 2001; Jerome et al. 2001; Lindsay, Vitelli et al. 2001; Merscher et al. 2001). While some evidence suggests that *Tbx1* is not expressed in cardiac neural crest cells, *Tbx1* expression is required for normal aortic arch development, as only a single aortic arch was present in *Tbx1* homozygous mutant cells (Lindsay and Baldini 1998). It is possible that *Tbx1* has the ability to trigger cellular signalling to induce artery development, affecting several molecules such as *Fgf8*. Again, a genetic link between *Tbx1* and *Fgf8* was demonstrated through an animal model: mice doubly heterozygous for *Tb1* and *Fgf8* had significantly more severe cardiovascular defects than single gene knockouts (Vitelli et al. 2002). While *Tbx1* may play a role in the 22q11DS, further analysis is needed to fully understand its contribution to the development of this disorder, as at least one patient has

been reported with the full spectrum of cardiac and craniofacial anomalies who carries an unusual 22q11DS deletion which does not involve *TBX1* (Yamagishi, Garg et al. 1999). A search for *TBX1* mutations in individuals with 22q11DS but with no detectable deletion or patients with congenital heart disease did not reveal obvious detrimental mutations, though the significance of several missense sequence changes remains to be determined (Lindsay, Vitelli et al. 2001). Further animals are also being designed in order to screen genetic modifiers of *Tbx1* as well as to better understand embryonic cardiac development. These critical animal studies of 22q11DS, combining large-scale deletion analysis as well as single candidate gene knockouts, are a model for the overall *in vivo* analysis of candidate genes in SMS.

Animal studies of AGS involving the mouse *Jag1* gene have also proved to be quite complex. As mentioned above, a family carrying a particular missense mutation in *JAG1* displayed hearing loss and certain vestibular abnormalities, as well as cardiac defects (Le Caignec, Lefevre et al. 2002). Vestibular defects were also recapitulated in two separate mouse lines carrying randomly generated *Jag1* missense mutations (Kiernan et al. 2001; Tsai et al. 2001). While human *JAG1* mutation screening supports the hypothesis that JAG1 haploinsufficiency is causative for AGS, gene targeting in mice to reduce *Jag1* expression revealed that homozygous gene-targeted (or knockout) mice died *in utero* and *Jag1/+* heterozygotes displayed some anterior chamber eye defects but otherwise were a disappointing phenotypic model for the other features of human AGS (Xue et al. 1999). This result is not unique to AGS, as several animal models for recessive metabolic disorders exists, which essentially do not recapitulate the human phenotype (Elsea et al. 2002). However, a separate mouse model was engineered that

combined a *Jag1* mutation as well as a *Notch2* (a member of the Notch family of transmembrane receptors that interacts with ligands such as *Jag1*) hypomorphic allele. These mice exhibited several of the phenotypic features similar to AGS patients such as jaundice, bile duct, heart, eye, and kidney abormalities (McCright et al. 2001) and enhanced the understanding of the *Notch* developmental pathway. Further animal studies are ongoing to understand the role of other genetic modifiers that may also influence the variable phenotypic expressivity seen in human AGS patients.

In order to analyze WS candidate genes, mice hemizygous for elastin (*Eln*) were also found to be clinically normal, though some histological evidence demonstrated increased elastin lemellae in the artery walls (Li et al. 1998). Mice homozygous for the *Eln* disruption showed increased thickness of the arterial wall due to smooth muscle accumulation after embryonic day 17.5 and died soon after birth (Li et al. 1998; Li, Faury et al. 1998). Similar to human patients who carry *ELN* mutations, mice hemizygous for *Eln* did not display obvious intellectual impairment (Li, Brooke et al. 1998). A possible candidate gene for the cognitive abnormalies found in WS, *Cyln2*, which encodes Clip-115, a microtubule-binding protein, was also removed in mice by gene-targeting (Hoogenraad et al. 2002). Mice hemizygous for *Cyln2* display some characteristics similar to WS patients, including growth deficiency, motor deficits, and several brain abnormalities, which were assessed with several sophisticated behavioral tests (Hoogenraad et al. 1998). As discussed in Chapter IV, we also intend to apply several behavioral analyses to discern subtle mental and behavioral defects in gene-targeted and transgenic mouse models for SMS. Animal models such as these are very important to

the identification of new cellular pathways which underlie complex neurobehavioral development.

Another form of mouse *in vivo* gene analysis, BAC transgenesis, will also be discussed in Chapter IV, and how the Elsea laboratory used this technology to determine the dosage sensitivity of one of the candidate genes in the SMS deletion region. While BAC transgenics do not appear to be widely used except for certain complementation experiments, they can be useful for preliminary transgenic studies without complex genetic engineering (Heintz 2001). In a classic example, the Camper lab introduced a BAC containing the unconventional myosin Myo15 gene into a line of *shaker-2* mutant mice (Probst et al. 1998). The *shaker-2* mice, whose phenotype included deafness and a characteristic circling behavior, were found to harbor a mutation within the motor domain of *Myo15*. The transgenic BAC supplied the wild type form of Myo15 and fully restored normal hearing and behavior to these mice (Probst, Fridell et al. 1998). In another example, a group studying DGS syndrome used human BACs overexpressing several human 22q11 candidate genes to assess which genes in the region produced measurable phenotypic effects when present in multiple copies (Funke et al. 2001; Merscher, Funke et al. 2001). The transgenic animals displayed severe inner ear defects and hyperactive behavior (Funke, Epstein et al. 2001). The expression pattern of *Tbx1* suggests that this gene may be responsible for the ear disorders in the BAC transgenic mice as well as the hearing loss present in some DGS patients (Funke, Epstein et al. 2001). As described in Chapter IV, we used a similar analysis to examine overexpression of mouse *Rai1*.

Conclusions

As highlighted in the examples above, there are several issues to consider when undertaking a genetic assessment in a live animal model such as mice, including strain differences, the lack of an obvious phenotype, and the role of genetic modifiers. In the development of mouse models for SMS candidate genes, we are also concerned about embryonic lethality. Several of the SMS candidate genes targeted in mice were found to be essential for embryonic development (Vlangos et al. 2003); we believe there is a possibility that true knockouts for many of the promising SMS candidate may die *in utero*. In that case, more sophisticated chromosomal engineering techniques utilizing the Cre and *lox*P recombination system may have to be applied to specifically remove a particular candidate gene at certain temporal stages, in order to properly assess the effect of dosage on the developing mouse embryo.

Chapter II.

Fine-mapping the SMS critical interval and identification of SMS candidate genes

The research goal in our lab is to understand the underlying molecular and biochemical basis of SMS, and the cellular basis of the complex physical and neurobehavioral phenotype manifested by SMS patients. In order to proceed effectively, we required a reliable physical and transcription map of the smallest deleted region along 17p11.2 which could still reproduce the entire SMS phenotype, or what is termed the critical interval or region. Previous to this work, a great deal of effort was put forth to identify the critical interval and to begin to map genes, ESTs, and genomic markers to the SMS region of 17p. A preliminary genomic contig was also created. The first major focus of my work was to complete the fine-mapping of the critical interval and enhance the transcription map, simultaneously identifying the genes and expressed transcripts which localize to this region. We would then prioritize these genes as possible SMS candidate genes.

Definition of the SMS critical interval

The definition of the SMS critical interval (published in 1997) was made possible by advances in FISH detection, which identified SMS patients with smaller or unusual 17p11.2 deletion (Juyal et al. 1996; Elsea et al. 1997). In order to further analyze these deletions, somatic rodent:human hybrid cell lines that retain the deleted copy of chromosome 17 were developed from 30 patients (Guzzetta et al. 1992; Elsea, Purandare

et al. 1997). These cell lines were created through a fusion and selection process in which an initial, single cell with two nuclei (human and rodent) is formed. After several rounds of division, human chromosomes are lost randomly. The hybrid eventually stabilizes and very few human chromosomes remain, making it possible to analyze the genes present on a single chromosome more efficiently (Kao et al. 1976). The SMS common deletion interval (typical for ~75% of SMS patients carrying deletions) is bordered by marker D17S58 (proximal) and cosmid cCI17-498 (distal) (Juyal, Figuera et al. 1996; Elsea, Purandare et al. 1997). The SMS patient with the smallest deletion detected to date, HOU142-540 represents the ~1.5 Mb SMS critical interval which lies between marker D17S29 (centromeric) and cosmid cCI17-638 (telomeric) (Elsea, Purandare et al. 1997). Immortalized cell lines from 2 non-SMS patients with overlapping deletions along 17p11.2-p12 were also created, HOU92-357 (Hy357-2D) and HOU261-765 (Hy765-18D), for further subdivision of the critical interval (Juyal, Figuera et al. 1996; Elsea, Purandare et al. 1997). Eventually, a 17p mapping panel was created that also included a partial 17p arm control (88H5), and an iso-17q (LS-1) (Elsea, Purandare et al. 1997).

Once a somatic cell hybrid mapping panel was established, the localization of numerous markers along chromosome 17p11-p12 was carried out, including short tandem repeats (STRs), sequence-tagged sites (STSs), expressed sequence tags (ESTs) and known genes (Elsea, Purandare et al. 1997). As of 1997, within the ~1.5 Mb critical interval, 23 genes/ESTs were mapped that were deleted in all patients. This was accomplished through polymerase chain reaction (PCR) fine mapping with primer sets

from various markers to DNA from the rodent:human somatic cell hybrid panel. The presence or absence of a marker in naturally-occurring deletions along 17p (from the hybrid panel) also demarked "bins," which divided the common and critical deletion regions. Additional mapping information was gained from restriction fragment length polymorphisms or FISH analysis with cosmids representing various genomic loci (Elsea, Purandare et al. 1997).

Physical and transcription map of the SMS critical interval

The definition of the SMS critical interval and a method to efficiently map markers within this critical region established a basis for the contiguous physical and transcription map of the SMS critical interval. After difficulty was encountered in attempting to reproduce a yeast artificial chromosome (YAC) map of the region (Chen, Manian et al. 1997), large insert genomic clones such as bacterial and P1-artificial chromosomes (BAC/PACs) were chosen to create the physical contig of the SMS critical interval. BACs and PACs within the contig were assembled through a combination of Southern hybridization, PCR, and human genome project sequence data analysis. Initially, known genes, ESTs, and genomic markers that mapped to the critical interval were used to screen a gridded hCITB BAC library in order to form a preliminary contig. Overlapping BACs were identified and grouped according to *Alu*-PCR fingerprinting and restriction digests. In order to map markers to the initial contig, known genes and ESTs were hybridized to *Eco*RI-digested BACs or PCR-amplified from BAC DNA. This preliminary contig had many gaps and could not be completed with the BAC resources that were available.

This work focuses on fine-mapping the SMS critical interval and completing the genomic contig, which contained several gaps in 1998. As late as 1999, little human genome project sequence data were available for 17p11.2, most likely due to the low-copy repeats that are prominent in this genomic region which can complicate sequence sequence alignment (Stankiewicz, Shaw et al. 2003). As soon as draft sequence for the unique, gene-rich regions of the SMS critical interval was available publically through NCBI, we were able to take advantage of these data to improve the coverage of our contig. We used BLAST searches with known 17p11.2 gene and EST sequence data to identify high-throughput genomic sequence (HTGS) RPCI BAC/PAC and CIT BAC data that were pertinent to the SMS critical interval. We subsequently purchased several of these clones for analysis and utilized draft sequence alignments from the Golden Path (http://genome.ucsc.edu) and Ensembl (http://www.ensembl.org) databases to organize our contig. We searched BAC-end sequence information from GenBank to establish a minimum tiling path of 16 BACs and 2 PACs (Figure 2) (Lucas et al. 2001). Several overlapping clones were also included in the map to provide greater coverage (Figure 2).

Gene order within the contig was determined through HTGS sequence analysis and Southern hybridization for the presence or absence of a certain gene or EST in EcoRI-digested BAC or PAC DNA. Transcription orientation was obtained from the human genome browser database, cosmid sequencing, and known overlapping segments, such as the 3' ends of FLII and LLGL1. Cosmid genomic clones for several of the genes were identified through hybridization of EST inserts to gridded filters of the Los Alamos

Figure 2. Transcription map of the ~1.5 Mb SMS critical interval (adapted from Lucas et al., 2001.

The SMS critical interval as of 2001 (originally determined in Juyal et al., 1996 and Elsea et al., 1997) spans between the two vertical hatch marks, approximately from the middle SMS-REP (SMS-REPM) to the distal SMS-REP (SMS-REPD). The transcription map is comprised of an overlapping contig of BACs and PACs from the proximal SMS-REPM region to the distal genomic marker D17S740. Markers, including ESTs and known genes, were mapped to the contig through a combination of hybridization to *EcoRI*-digested BAC and PAC DNA, PCR, and database sequence analysis from the human genome project. All ESTs and genes were also hybridized to the somatic cell hybrid mapping panel to confirm HTGS sequence analysis and position within the SMS critical interval. BAC/PAC-end sequence overlaps confirmed the position of clones where markers are absent. Genes, ESTs, and markers shown in blue are localized out of the current SMS critical interval.

32
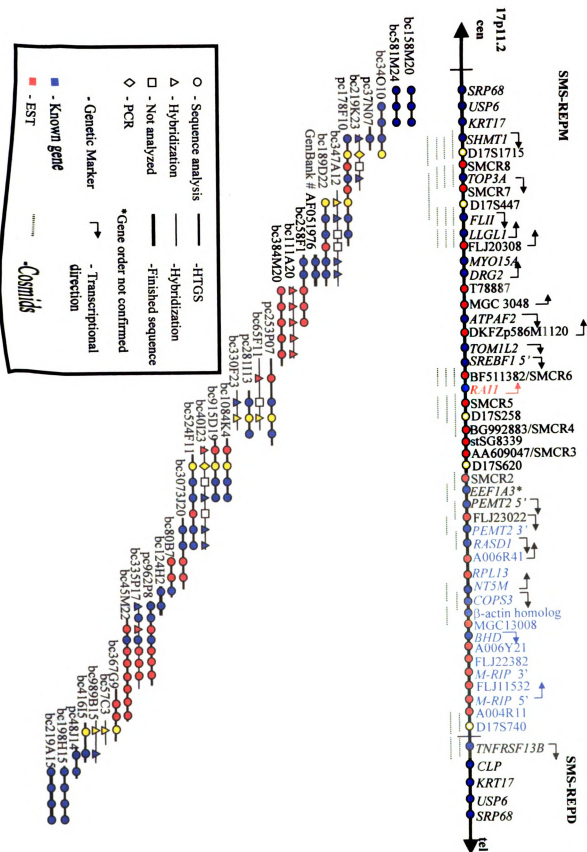
flow-sorte

from SH

order. We

hybridizin

DNA (Luc

representi

(Lucas, V

control hy

As

SMS patie

Vlangos et

within the

PACs DN

cDNA clo

membrane

contig, as

published c

were utilize

D17S740, D

1041) (Luca

flow-sorted chromosome 17-specific cosmid library (Kallioniemi et al. 1994). The region from *SHMT1* to *FLII* is well-represented by overlapping cosmids, which confirmed gene order. We reconciled human genome project data with naturally-occurring deletions by hybridizing genes and ESTs to *Eco*RI-digested rodent:human somatic cell hybrid panel DNA (Lucas, Vlangos et al. 2001). An example is shown in Figure 3, in which an EST representing the 3' end of *RAI1* is hybridized to this somatic cell hybrid mapping panel (Lucas, Vlangos et al. 2001). *RAI1* is deleted in all SMS patients but is present in the control hybrid lines, MH22-6 and 88H5, as well as human genomic DNA.

As published in 2001, the ~1.5 Mb SMS critical interval is represented by the SMS patient HOU142-540, who carries the smallest deletion described to date (Lucas, Vlangos et al. 2001). All ESTs were also hybridized to *Eco*RI-digested BACs and PACs within the contig to confirm mapping positions. In Figure 4, the digested BACs and PACs DNA from within the SMS critical interval is shown in (a), and in (b), an ~2.0 kb cDNA clone containing the 3' end of *RAI1* demonstrates positive hybridization to a membrane containing the digested BAC/PAC DNA. *RAI1* maps to pc253P07 within the contig, as predicted from human genome project draft sequence. Within this 2001 published contig, 17 known genes and 12 ESTs were identified and six genomic markers were utilized in the generation of the physical map of the critical region: D17S258, D17S740, D17S1794, D17S620, D17S447 (also called FG2), and D17S71 (also called A-1041) (Lucas, Vlangos et al. 2001).

Figure 3. Mapping *RAI1* to somatic cell hybrids.

In order to fine-map *RAI1* to a specific region within 17p11.2, the ~2.0 kb insert from EST DKFZp434A139Q2 was hybridized to a panel of *Eco*RI-digested somatic cell hybrids, carrying deleted chromosomes 17 from SMS patients (Hy147-20D, Hy540-1D and Hy484-2D), and non-SMS patients with overlapping deletions along 17p11.2 (Hy765-18D and Hy357-2D) as well as chromosome 17p positive controls (MH22-6, human genomic, and Hy88H5), and 17q (LS-1), hamster (a23), and mouse (Cl-1D) cell line negative controls. A >20 kb *RAI1* band is evident only in the MH22-6, human, Hy88H5, and Hy357-2D lanes (present but faint on this exposure), demonstrating that *RAI1* is deleted in all SMS patients and maps to the central portion of the SMS critical interval between D17S258 and D17S620.

MH22-6

Human genomic

Hy88H5

Hy147-20D

Hy540-1D

Hy765-18D

Hy357-2D

Hy484-2D

LS-1

Hamster

Mouse

17p

>20 kb →

11.2

12

13.1

13.2

13.3
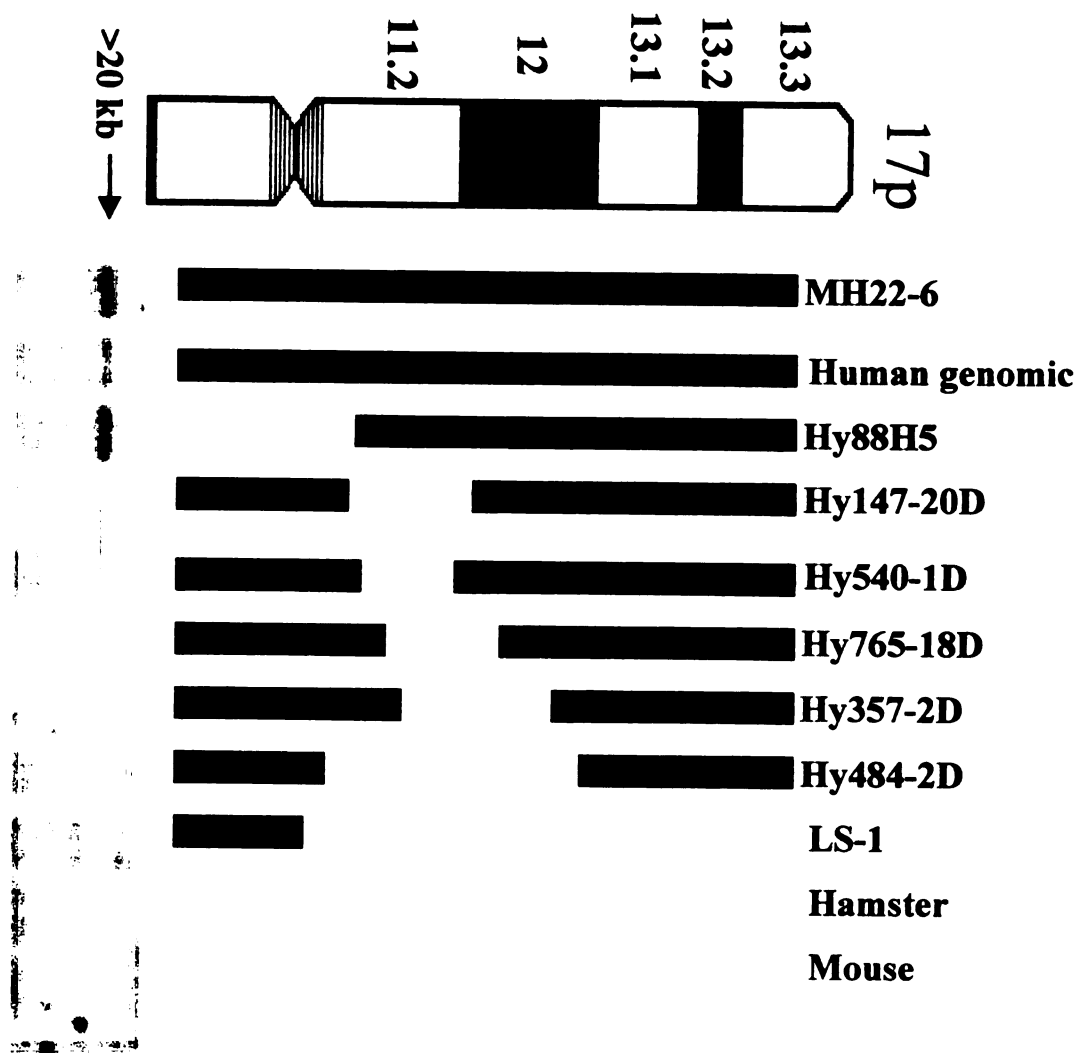
Figure 4. Mapping _RAI1_ to BACs/PACs.

(a) DNA (8 µg) from BACs within the SMS critical interval and human genomic DNA (10 µg) were digested with _Eco_RI and electrophoresed overnight. The digested DNA was transferred to a nylon membrane using standard Southern techniques.

(b) An autoradiograph of the Southern blot is shown, hybridized to the DKFZp434A139Q2 EST plasmid insert, which represents the 3' end of human _RAI1_. Red arrows indicate positive _RAI1_ hybridization in pc253P07 (which confirms human genome project sequence analysis) and a faint band present in the human lane. BAC-end hybridization is seen in several other lanes.
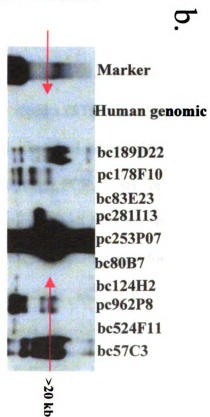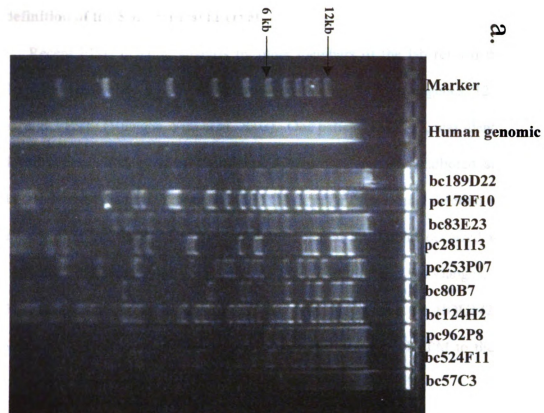
Re-definition of the SMS critical interval

Recent FISH deletion analysis by other members of the lab refined the current SMS critical interval to a ~950 kb interval, containing 13 known genes, 12 predicted genes, and 3 ESTs (Vlangos, Yim et al. 2003). One patient, SMS135, was identified who displayed typical SMS physical features and behaviors but who harbored an unusual 17p11.2 deletion (Vlangos, Yim et al. 2003). This patient's proximal deletion breakpoint occurred within the SMS-REPP but the distal deletion breakpoint occurred within the ~1.5 Mb SMS critical interval, proximal to *RASD1* (Figures 2 and 5) (Vlangos, Yim et al. 2003). Based on our lab's analysis of the deletions carried by HOU142-540 and SMS135, the current SMS critical interval spans from the SMS-REPM to the genomic region just proximal to *RASD1* (Figure 5) (Vlangos, Yim et al. 2003).

EST characterization and analysis

To define the ~1.5 Mb SMS critical interval, numerous markers were mapped along chromosome 17p11-p12, including short tandem repeats (STRs), sequence-tagged sites (STSs), expressed sequence tags (ESTs), and 12 identified genes. Binning/fine-mapping of ESTs along 17p11.2 was accomplished through PCR, FISH, or Southern hybridization to DNA from a panel of rodent:human somatic cell hybrids. In order to more precisely map these markers to the transcription map, plasmid clones for 13 ESTs that mapped to the critical interval were obtained commercially, and DNA was isolated. All clone inserts were sequenced and extensive database searches of the nucleotide sequence and 6-frame amino acid translation were conducted in order to identify homologous genes and sequence motifs. We determined the tissue expression pattern

Figure 5. SMS deletion analysis and refined SMS critical interval (adapted from Vlangos et al., 2003).

The ~1.5 Mb SMS critical interval, which was mapped in Lucas et al, 2001, is defined by the deletion breakpoints from SMS patient HOU142-540. New deletion analysis from patient SMS135 re-defines the distal end of the ~950 kb SMS critical interval, which extends from SMS-REPM to the genomic region just proximal to *RASD1*. Relevant genomic markers are included, which map to SMS critical interval as well as the ~4 Mb SMS proximal and distal common interval.

Proximal Common Interval   Distal Common Interval

~950 kb   ~1.5 Mb   ~4 Mb

SMS-REPP   D17S805   D17S29   SMS-REPM   D17S1715   D17S740   SMS-REPD

Common Interval

HOU142-540

SMS135

Critical Interval

40
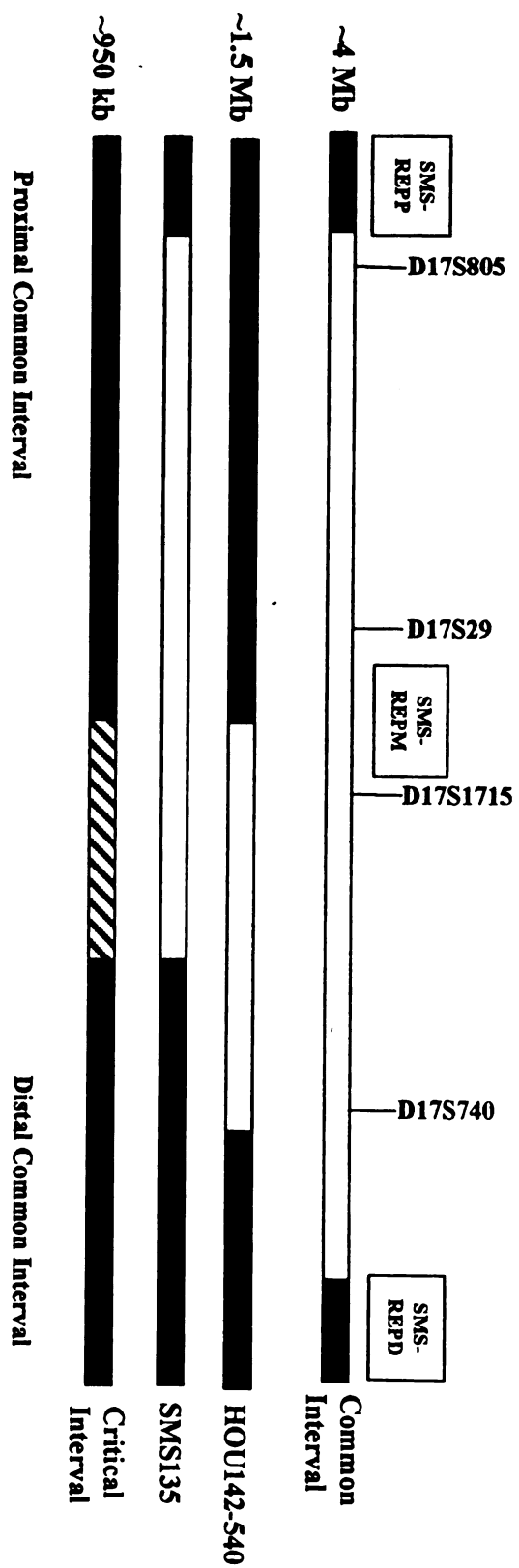
Figure 6. Human *RAI1* northern analysis.

The ~2.0 kb insert from EST DKFZp434A139Q2, representing the 3' end of human *RAI1*, was hybridized to an adult Clontech MTN northern blot (shown in the panels on the left) and a fetal Clontech northern blot (shown in the panels on the right). An ~8.0 kb transcript is evident in all adult and fetal tissues. β–actin is included as a mRNA loading control for both blots.

Top panel:
- β-Actin lane markers: 1.35, 2.4
- RAII lane markers (kb): 7.5, 9.5
- Tissue labels (top to bottom): Heart, Brain, Placenta, Lung, Liver, Sk. muscle, Kidney, Pancreas

Bottom panel:
- β-Actin lane markers: 1.35, 2.4
- RAII lane markers (kb): 7.5, 9.5
- Tissue labels (top to bottom): Brain, Lung, Liver, Kidney

42
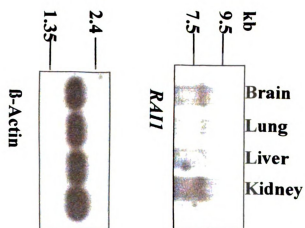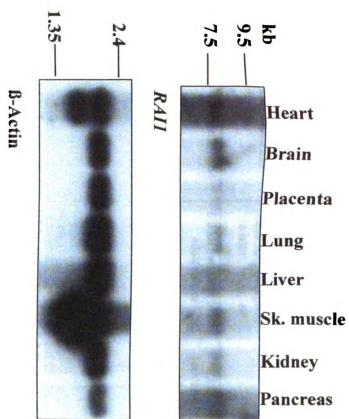
EST amplimer(s) (GenBank accession)	Unigene	Tissue expression	Protein motifs	Known gene
WI-13499 (GenBank R41366)	Hs.78582	2.2 kb transcript in all adult and fetal tissues[a]	GTP-binding protein motif	DRG2 (GenBank X80754)
D17S2021 (GenBank AA281720)	Hs.13434	1.8 kb transcript in all adult and fetal tissues; highest in heart	Mitochondrial leader sequence	ATP4F2 (GenBank AF052185)
A003A44 (GenBank AA236905) stSG9692 (GenBank AF038192)	Hs.8125 Hs.12537	6.5 kb transcript in all adult tissues, highest in heart, brain, sk. muscle	VHS and GAT domains	TOM1L2 (GenBank NM_144678)
DKFZp434A13902 (GenBank AL133649)	Hs.278684	8.0 kb transcript in all adult tissues	CAG repeat, nuclear localization signal, PHD zinc finger domain	R4I1 (GenBank AY172136)
NIB1041 (GenBank T16275)	Hs.106359	5 kb transcript in all adult tissues[a]	Member of Ras subfamily	RASD1 (GenBank NM_016084)
WI-11472 (GenBank R72633)	Hs.16614	1.6 kb transcript in adult heart, brain, and skeletal muscle	Hydrolase domain	NT5M (GenBank NM_020201)
stSG526124 (GenBank AI570799)	Hs.34883	6.0 kb transcript in all adult and fetal tissues	KOG4807: domain from F-actin binding protein	M-RIP (interim gene symbol) (GenBank NM_015134)

Table 1. ESTs mapped by Lucas et al. within the 2001 ~1.5 Mb critical interval which have been identified as known genes. Genes have been ordered proximal to distal and shading indicates gene is located outside of the current SMS critical interval (Vlangos et al., 2003). Protein motifs identified from NCBI conserved domain database and NCBI protein-protein BLAST alignments.
[a]Results reported in Vlangos et al., 2000; [b]Northern results similar to those reported in Rampazzo et al., 2000; [c]Northern results similar to those reported in Tu and Wu, 1999

remains a possibility that some of these unique ESTs can contains genomic contamination or may be cloning artifacts. If protein motifs were identified from NCBI conserved domain database and NCBI protein-protein BLAST alignments.

EST Amplifier (GenBank Accession)	Unigene	Tissue expression	Protein motifs
T78887 (GenBank AA730163)	Hs.187422	Clontech MTE blot showed expression in heart, GI tract, germ cell, lung, kidney, placenta	378 bp LINE1 repeat
ssGC8339 (GenBank H57290)		Clontech MTE blot showed faint expression in heart, GI tract, brain, lung, placenta, germ cell	93 bp MIR repeat
FL20308 (GenBank BC062339)	Hs.356770	3.5 and 4.0 kb transcript in all adult and fetal tissues examined	KOG4176: uncharacterized conserved domain
MGC3048 (GenBank BC000636)	Hs.11437	4.4 kb and 4.0 transcript in adult heart, sk. muscle, kidney, pancreas and fetal liver and lung; 4.0 kb transcript in adult and fetal brain, lung and placenta	KOG4635: domain found in vacuolar import and degradation proteins
DKFZp586M1120 (GenBank AL136926)	Hs.159068	Unigene reports 5.8 kb transcript and ESTs from several adult tissues	43% similarity (140 aa alignment) to sds22 protein homolog (GenProt S68209)
FLJ23022 (GenBank NM_025051)	Hs.287717	Unigene reports SAGE tags from several adult tissues	none
A006121 (GenBank H50830)	Hs.31652	Unigene reports ESTs from several adult tissues	322 bp Alu repeat, 85.7% alignment to phospholipase D active site motif
FLJ22382 (GenBank AK026035)	Hs.46783	ESTs from several adult tissues	none
FLJ11532 (GenBank AK021594)	Hs.296656	ESTs from brain and embryo	78 bp Alu repeat
A044R1T (GenBank T91728)	Hs.16899	2.4 kb transcript in adult and fetal liver	none
A006R41/FLJ10193 (GenBank AI500641)	Hs.235195	6.0 kb transcript in adult skeletal muscle and fetal brain	Moderately similar to Drosophila protein RH42446p (GenProt AAN71576)
MGC13008 (GenBank NM_032680)	Hs.326732	ESTs from rhabdomyosarcoma	none

Table 2. Novel ESTs mapped by the Elsea lab within the 2001 ~1.5 Mb critical interval. Genes have been ordered proximal to distal and shading indicates that this gene is localized outside of the current SMS critical interval (2003). It remains a possibility that some of these unique ESTs can contains genomic contamination or may be cloning artifacts. Protein motifs were identified from NCBI conserved domain database and NCBI protein-protein BLAST alignments.

EST Amplimer (GenBank Accession)	Unigene	Tissue Expression	Protein motifs	References
SMCR8 (GenBank AF467440)	Hs.513986	2.8, 3.0, and 6.5 kb transcript in all adult and fetal tissues	KOG3715: domain found in LST7 permease Golgi transport protein	Bi et al. (2002)
SMCR7 (GenBank AF467443)	Hs.100448	2.4 and 3.4 kb transcripts in all adult and fetal tissues, predominantly in heart and sk. muscle	KOG3963: motif found in Mab-21-like cell fate proteins	Bi et al. (2002)
SMCR6 (GenBank BF511382)	Hs.443639	RT-PCR demonstrated expression in all tissues	none	Bi et al. (2002)
SMCR5 (GenBank AF467442)	Hs.352643	RT-PCR demonstrated expression in all tissues	none	Bi et al. (2002)
SMCR4 (GenBank BG992883)	none	RT-PCR demonstrated expression in all adult tissues and fetus	none	Bi et al. (2002)
SMCR3 (GenBank AA609047)	Hs.373802	RT-PCR demonstrated expression in all adult tissues and fetus	none	Bi et al. (2002)
SMCR2	none	RT-PCR demonstrated expression in adult brain, and fetus	none	Bi et al. (2002)
SMCR9	none	RT-PCR demonstrated expression in all adult tissues and fetus	Two PDZ domains	Bi et al. (2002)
BHD (GenBank AF517523)	Hs.396533	3.8 kb transcript in all adult tissues	KOG3715: domain found in LST7 aa permease Golgi transport protein	Nickerson et al. (2002)

Table 3. ESTs recently mapped by other groups within the 2001 ~1.5 Mb critical interval. Following the publication of our physical and transcription map in 2001 (Lucas et al., 2001), several other genes were mapped to this region by other groups. Genes have been ordered proximal to distal and shading indicates that this gene is localized outside of the current SMS critical interval (Vlangos et al., 2003). Domains were identified from NCBI conserved domain database.

and tran

Figure

expressi

RA11 ge

is the ex

the SMS

now bec

ESTs th

novel E

data ab

been ma

Table

unident

be clon

project

databas

interval

evidenc

role in

the cri

and transcript size for each EST by northern analysis of adult human and *fetal mRNAs*. Figure 6 shows the multiple tissue northern (MTN) and human *fetal tissue blot* expression pattern of EST DKFZp434A139Q2, which represents the 3' end of the human *RAI1* gene. An ~8.0 kb transcript is observed in all adult and fetal tissues examined; this is the expected *RAI1* mRNA size for the major transcript (Toulouse et al. 2003).

Tables 1-3 show the current status of the genes and ESTs which were mapped to the SMS critical interval in 2001. Several transcripts, shaded in grey in each table, have now been localized outside of the current SMS critical region. Table 1 highlights seven ESTs that we mapped to the contig in 2001, which represent known genes. Several other novel ESTs were mapped and characterized by our group; current expression and protein data about these transcripts is presented in Table 2. Table 3 describes genes that have been mapped to the ~1.5 Mb SMS critical interval by other research groups. As shown in Table 2, several of the ESTs that we sequenced were novel or only matched other unidentified clones in the database. It remains a possibility that some of these ESTs may be cloning artifacts or may contain genomic contamination. As the human genome project and large-scale EST projects continue to deposit more information in the public databases, a number of transcripts continue to be identified within the SMS critical interval and expression data for the novel genes and ESTs increases. Experimental evidence will be needed to determine if any of these novel genes may play a phenotypic role in SMS. And while we have focused our research efforts on genes and ESTs within the critical interval, it remains a possibility that 17p11.2 deletion breakpoints alter

regulatory and binding elements, which can affect the expression of genes outside the deleted region. This phenomenon could contribute to the SMS phenotype.

Summary and analysis of candidate genes

Once our physical and transcription map of the SMS critical interval was published, we began to prioritize the genes within this region as candidates for the SMS phenotype. Several characteristics that we considered when identifying SMS candidate genes were: expression in brain, developmental expression, known protein motifs, DNA-binding domains, and protein interaction domains. Following the discovery that haploinsufficiency of the transcription factor *Tbx1* may play an important role in 22q11DS, we were especially interested in putative transcription factors which may have a global impact on embryonic development, though other signalling molecules could also prove to be especially dosage-sensitive. However, many of genes that mapped to the SMS critical interval were novel and showed expression pattern in every tissue analyzed (Tables 1-3). We relied on other factors to reduce the number of candidate genes that we considered in our studies. To date, genes distal to *PEMT2* are no longer considered part of the current SMS critical interval and are therefore a low priority for future analysis as SMS candidate genes (Figure 2). Several of the known genes that map to the critical interval including, *TOP3A* (Hanai et al. 1996; Fritz et al. 1997; Elsea et al. 1998), *FLII* (Chen et al. 1995; Campbell et al. 1997; Campbell et al. 2000; Campbell et al. 2002), *MYO15A* (Wang et al. 1998; Liang et al. 1999), *SREBF1* (Elsea, Purandare et al. 1997), *PEMT2* (Walkey et al. 1999), and *COPS3* (Elsea et al. 1999; Potocki et al. 1999; Yan et al. 2003) have been well-characterized, and we also do not consider them priority

Mouse gene disrupted	Heterozygous phenotype	Homozygous phenotype	Reference
Top3a	No abnormalities	Embryonic lethal	Li and Wang (1998)
Flii/h	No abnormalities	Embryonic lethal	Campbell et al (2002)
Myo15	No abnormalities	Profound deafness and vestibular defects	Probst et al. (1998)
Srebf1	No abnormalities	No abnormalities in normal births; ~65% embryonic lethal	Shimano et al. (1997)
Pemt	No abnormalities	No abnormalities on normal diet	Walkey et al. (1997)
Cops3	No abnormalities	Embryonic lethal	Yan et al. (2003)

Table 4. Phenotype of SMS candidate genes disrupted in mouse. This table was adapted from Vlangos et al. (2003).

candidates genes for SMS. As described in Table 4, several genes homologous *to SMS* candidate genes, such as *Top3a, Fliih, Myo15, Srebf1, Pemt,* and *Cops3* have been analyzed in gene-targeted mice and in all cases heterozygous animals appeared normal and healthy, which therefore suggests that these genes are not particularly dosage-sensitive. In the case of *MYO15A*, mutations in this gene are causative for hereditary recessive deafness (DFNB3), and this gene was only found to contribute to the deafness of one SMS patient who harbored a *MYO15A* missense mutation on this individual's non-deleted chromosome 17 (Liburd et al. 2001). The SHMT1 protein has also been analyzed in patient lymphoblastoid cell lines and while haploinsufficiency of SHMT1 was detected, serum folate, glycine, and serine levels were normal (Elsea et al. 1995). We have not completely disregarded SHMT1 as an SMS candidate gene, as studies in other types of patient cell lines or cerebrospinal fluid may be more conclusive.

There are several known genes within the current SMS critical interval which have not been extensively analyzed and that we consider to be possible SMS candidate genes based on expression data and their putative cellular role. A summary of these 5 genes, listed proximally to distally, is given below:

LLGL1: *LLGL1* (GenBank NM_004140; Unigene Hs. 95659) is the human homolog of the *Drosophila* tumor suppressor gene, *D-lethal (2) giant larvae*. When disrupted, *D-lgl* can produce tumors in imaginal disks and abnormal transformation of the adult optic centers in larval brains (Strand et al. 1995). Human *LLGL1* is expressed in brain, kidney, and muscle, and shows an association with the cytoskeleton (Strand, Unger et al. 1995). Antibodies against *LLGL1* were able to co-immunoprecipitate *LLGL1* and non-muscle

myosin heavy chain; this interaction is conserved between human and *Drosophila* (Strand, Unger et al. 1995). Both *Drosophila* and human LLGL1 are also associated with a serine kinase, which is able to specifically recognize and phosphorylate serine residues within the protein (Strand, Unger et al. 1995). The *LLGL1* cDNA encodes a protein of 1057 aa, with a predicted molecular weight of 115 kDa (Strand, Unger et al. 1995). *LLGL1* was mapped to chromosome 17p11.2 by FISH and Southern blotting (Strand, Unger et al. 1995), and was demonstrated to overlap at the 3' end with the *FLII* gene (Campbell, Fountain et al. 1997). *LLGL1* is transcribed in the opposite orientation as *FLII*, just telomeric to *FLII* in the SMS critical interval, and is found on bc347A12, bc189D22, and pc178F10 (Figure 2). The contribution of this gene to the SMS phenotype is unknown.

DRG2: Directly adjacent to *MYO15A* is developmentally-regulated GTP-binding protein 2 *(DRG2)* (GenBank NM_001388; Unigene Hs.78582) (Liang et al. 1999). Fine mapping of this gene, represented by the EST WI-13499, confirms that *DRG2* maps to 17p11.2 (Vlangos et al. 2000), within the SMS critical region. *DRG2* maps to bc347A12 and bc189D22 (Figure 2), as well as a region of genomic sequence that was generated through *MYO15* analysis (GenBank AF051976). The DRG2 protein is a member of a novel family of GTP-binding proteins that are highly conserved and may play a role in oncogenesis (Schenker et al. 1994). *DRG2* is highly homologous to *DRG1*, a developmentally regulated gene that is expressed in embryonic and neoplastic tissue and down-regulated in adult tissues (Sazuka et al. 1992; Sazuka et al. 1992). Northern analysis of *DRG2* shows low levels of expression in all adult and fetal tissues examined

(Schenker, Lach et al. 1994; Vlangos, Das et al. 2000). Recently, cloned *Xenopus laevis*
drg1 and *drg2* indicates that the recombinant XDRG1 and XDRG2 proteins both have *in
vitro* RNA binding ability (Ishikawa et al. 2003). The *in vivo* cellular role of this
extremely well-conserved gene is currently unknown.

***ATPAF2* (formerly *ATP12*):** ATPAF2 is the human homolog of the yeast nuclear gene
Atp12p, which is required to mediate the formation of the F1$_\alpha$ subunit of the
mitochondrial ATPase (Wang et al. 2001; Bi, Yan et al. 2002). The cellular role of yeast
Atp12p appears to be limited to ATP synthase assembly (Wang, White et al. 2001). We
showed that the EST D17S2021, which represent *ATPAF2*, is expressed on a northern
blot in all tissues examined (Table 1); further studies using quantitative reverse-
transcription PCR indicated that expression of mouse *Atpaf2* mRNA varied up to 30-fold.
The highest mRNA expression was found in brown adipose tissue (Pickova et al. 2003).
The human ATPAF2 sequence contains a mitochodrial targeting function and a human
ATPAF2 cDNA has the ability to rescue a yeast *atpaf2* mutant, demonstrating an
equivalent cellular function (Wang, White et al. 2001; Ackerman 2002). The effect of
ATPAF2 haploinsufficiency in SMS patients is not yet known, though it is possible that
the hypotonia demonstrated in SMS patients could be related to abnormal mitochondrial
function mediated by ATPAF2 haploinsufficiency.

***TOM1L2*:** Our group mapped the target of myb1-like 2 (*TOM1L2*) ESTs A003A44 and
stSG9692 to the SMS critical interval (Figure 2 and Table 1), and the full-length human
and mouse transcripts of *TOM1L2/Tom1l2*, were subsequently assembled by the Lupski

group and determined to share a similar genomic structure (Bi, Yan et al. 2002). Little is known about the TOM1L2 protein, except that VHS and GAT domains are present, which have been implicated in intracellular trafficking and sorting (Bi, Yan et al. 2002). *TOM1L2* appears to be expressed ubiquitously, though at higher levels in heart, brain, and skeletal muscle (Table 1). The VHS domain, which is named for three hallmark proteins which contain this domain, Vps27, Hrs, and STAM (Lohi et al. 1998; Lohi et al. 2002) is found in several protein complexes which can interact to bind ubiquitin and sort ubiquinated proteins through the endosome (Yamakami et al. 2003). The GAT domain is found in proteins which have been implicated in sorting receptors through the *trans*-Golgi network (Shiba et al. 2003; Yamakami, Yoshimori et al. 2003). *TOM1L2* was named for its similarity to the *TOM1* gene, a specific target for the v-myb oncogene (Yamakami, Yoshimori et al. 2003). While Tom1 was recently shown to bind directly to ubiquitin and to interact with the Tollip protein in a yeast two-hybrid screen (Yamakami, Yoshimori et al. 2003), the specific cellular role of TOM1L2 is unknown. The Elsea lab is currently performing *in vitro* and *in vivo* analysis of *TOM1l2/Tom1l2* in order to understand whether cellular trafficking may be disrupted in some SMS patients.

RAI1: The identification and characterization of *RAI1* will be discussed at length in the following chapter.

Known genes originally mapped to the SMS critical interval

Three EST transcripts, NIB1041, WI-11472, and stSG26124 which were mapped to the ~1.5 Mb SMS critical interval contig have subsequently been recognized by other groups to be known genes, *RASD1*, *NT5M*, and *M-RIP*, respectively (Table 1). Though

these genes now map out of the current SMS critical interval, the cellular role of these genes is under investigation by other research groups and may prove to play a role in the modulation of the SMS phenotype.

RASD1 (formerly AGS1/DEXRAS1): We previously localized the EST marker NIB1041 to the SMS critical interval by PCR to somatic cell hybrids and by Southern blotting to BAC DNA (Elsea, Purandare et al. 1997). NIB1041 maps to the distal end of the 2001 critical interval and to bc524F11, bc3073J20, and bc40I23 in our transcription map (Figure 2). Sequence analysis in our laboratory has determined that the ~1.5 NIB1041 insert is a 99% match to the cDNA for activator of G-signaling 1 (*AGS1*) (GenBank AF069506, Unigene Hs.25829). The 1740 bp cDNA sequence for *AGS1* also matches the cDNA sequence for dexamethasone-induced ras-related protein (*DEXRAS1*) (GenBank AF262018, Unigene Hs.25829), although the complete *DEXRAS1* cDNA sequence in GenBank is 5141 bp long. Unigene has grouped *AGS1* and *DEXRAS1* together in the same entry, indicating that these sequences represent the same gene. *AGS1* (activator of G-protein signaling), was isolated through a *Saccharomyces cerevisiae* screen to identify mammalian nonreceptor activators of G-protein signaling pathways (Cismowski et al. 1999). AGS1, a ras-related protein, is able to process guanosine triphosphate exchange of the heterotrimeric G protein G_α subunit (Cismowski, Takesono et al. 1999; Cismowski et al. 2000). *DexRas1* was discovered independently in mice through differential display analysis to identify genes that were induced by glucocorticoid hormone (dexamethasone) treatment in AtT-20 cells (Kemppainen et al. 1998). Murine *DexRas1* is expressed in heart, brain, and liver, and *DEXRAS1* mRNA

levels in these tissues significantly increased after glucocorticoid treatment (Kemppainen and Behrend 1998). Human *DEXRAS1* cDNA was isolated through a yeast two-hybrid screen and subsequently cloned (Tu et al. 1999). The mouse and human DEXRAS1 protein are 98% identical at the amino acid level and both are members of the Ras superfamily of GTPases (Tu and Wu 1999). Expression of human *DEXRAS1* was also shown to be strongly stimulated by dexamethasone, in a variety of tissues (Tu and Wu 1999). Northern analysis demonstrated *DEXRAS1* expression in all adult tissues examined (Tu and Wu 1999). This result is in agreement with our northern analysis of EST NIB1041 (Table 1). Recent investigation into the role of rat Dexras1 in signaling demonstrates an interaction between the neuronal nitric acid (NO) synthase adaptor protein CAPON and Dexras1 to enhance NO signaling (Fang et al. 2000). Human *DEXRAS1* may also play a role in cell adhesion and extracellular matrix interactions, (Tu and Wu 1999) though the contribution to the SMS phenotype is unknown. In 2001, the Human Gene Nomenclature Committee formally changed the name of *AGS1/DEXRAS1* to *RASD1*. Recent work on *RASD1* has identified a functional glucocorticoid response element in the 3' untranslated region of the human *RASD1* gene (Kemppainen et al. 2003) and demonstrated that mouse *Rasd1* is expressed in a circadian manner in the suprachiasmatic nucleus of the brain (Takahashi et al. 2003). The latter finding may be relevant to SMS research because of the inverted circadian rhythm seen in SMS patients, though this gene has recently been mapped out of the SMS critical interval and may not be responsible for the majority of sleep abnormalities in SMS patients.

NT5M (dNT-2): Our studies mapped the EST WI-11472 to the distal region of the 1997

critical region, adjacent to *COPS3* (Figure 2) (Elsea, Purandare et al. 1997). BLAST

searches revealed that that the sequence of the ~2.1 kb WI-11472 insert is a 99% match

to the cDNA for deoxyribonucleotidase-2 *(dNT-2/NT5M)* (GenBank NM_020201,

Unigene Hs.16614), which codes for a mitochondrial deoxyribonucleotidase (Rampazzo

et al. 2000). Independently, Rampazzo et al. mapped *NT5M* to the SMS region based on

sequence analysis of a genomic clone (GenBank AC006071) and the proximity of *NT5M*

to *COPS3*, which is transcribed in the opposite orientation (Rampazzo, Gallinaro et al.

2000). Deoxyribonucleotidases hydrolyze the monophosphate ester linkages of the 5' or

3' carbon of deoxyribonucleotides, producing an inorganic phosphate and the

corresponding deoxyribonucleoside (Paglia et al. 1984). The full-length cDNA for

NT5M is 1617 bp long and contains a N-terminal mitochondrial leader sequence

(Rampazzo, Gallinaro et al. 2000). Northern analysis of *NT5M* revealed high expression

in the heart, brain, and skeletal muscle, a "typical" pattern for a mitochrondrial enzyme

(Rampazzo, Gallinaro et al. 2000). These data are in agreement with our northern

analysis of EST WI-11472 (Table 1). The authors demonstrated that NT5M is targeted to

the mitochondria by GFP-fusion localization and by *in vitro* incubation of the

recombinant protein with mitochondria (Rampazzo, Gallinaro et al. 2000). Excluding the

mitochondrial leader sequence, *NT5M* shares a 52% amino acid identity to *dNT-1*, a

ubiquitous, cytoplasmic deoxyribonucleotidase (Rampazzo, Gallinaro et al. 2000).

NT5M activity is limited to the dephosphorylation of thymine and uracil

deoxyribonucleotides and may play an important role in modulation of dTTP substrate

pools and removal of excess dTTP from the mitochondria (Rampazzo, Gallinaro et al.

2000). A potential role in SMS is yet to be identified, though, similar to *ATPAF2*, modulation of mitochondrial function by *NT5M* may be affected in SMS patients.

M-RIP: The myosin phosphatase-RhoA interacting protein, or *M-RIP* was recently identified through a yeast two-hybrid screen for proteins which interact with myosin phosphatase and RhoA (Surks et al. 2003). Our group mapped an EST for *M-RIP*, stSG26124, to the very distal end of the 2001 SMS critical interval, near the SMS-REPD (Figure 2) and demonstrated ubiquitous expression on a northern blot (Table 1). The human M-RIP is highly homologous to the murine p116RIP3 protein and was localized to actin myofilaments (Surks, Richards et al. 2003). M-RIP directly binds the myosin subunit of myosin phosphatase in smooth muscle. An adjacent domain of M-RIP binds RhoA, and these proteins may act to regulate myosin phosphatase (Surks, Richards et al. 2003), though currently no role in SMS has been proposed for this gene.

Conclusions

The physical and transcription map of the SMS critical interval and simultaneous gene identification highlighted many potential SMS candidate genes. Subsequent studies in our lab will focus on prioritizing the genes which may play a role in the SMS phenotype as well as in-depth analysis of expression pattern and determination of the *in vivo* cellular role of each potential candidate gene.

Materials and Methods

DNA Preparation:

BAC and PAC DNA extraction: BAC and PAC clones were obtained from Research Genetics (now Invitrogen) and BAC/PAC Resources to construct the SMS 17p11.2 contig, and DNA was isolated using a modified Qiagen very low copy protocol devised by our lab. Bacteria containing BACs or PACs were streaked on selective plates. Single colonies were picked and grown while shaking for 16 to 18 hours at 37°C in 2 ml of Luria-Bertani Broth (LB) containing selective antibiotic (12.5 μg/ml chloramphenicol for BACs and 100 μg/ml kanamycin for PACs). Starter cultures were then diluted 1/500 in 250 ml LB containing selective antibiotics and grown for 16 to 18 hours at 37°C while shaking. Bacteria were pelleted by spinning at 6000 x g for 15 minutes at 4°C. The supernatant was removed, and the pellet was completely resuspended in 60 ml of Buffer P1 (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0, 0.1 mg/ml RNase A). Sixty milliliters of Buffer P2 (0.2 N NaOH, 1% sodium dodecyl sulfate {SDS}) were then added to lyse the cells followed by 60 ml of Buffer P3 (3 M potassium acetate). The tubes were gently inverted and then placed on ice for 30 minutes to precipitate cellular debris. Following this incubation, the tubes were centrifuged at 4°C at ≥ 20,000 x g. Supernatant containing plasmid DNA was promptly filtered through filter paper (Whatman 4). BAC/PAC DNA was precipitated by adding 0.7 volumes of room temperature isopropanol and centrifuged at ≥ 15,000 x g for 30 minutes at 4°C. The pelleted DNA was allowed to dry slighty (inverted for 5 minutes) and then resuspended in 500 μl sterile 1x TE buffer (10 mM Tris-HCL, pH 8.0, 1 mM EDTA). After the DNA was resuspended, 5 ml of QBT (0.75 M NaCl and 50 mM MOPS buffer pH adjusted to 7.0,

containing 15% isopropanol and 0.15% Triton X-100) was added to the plasmid DNA and then the DNA was applied to an equilibrated Qiagen Tip 500. The column was washed 3 times with 10 ml of buffer QC (1 M NaCl, 50 mM MOPS buffer pH 7.0, containing 15% isopropanol). The BAC/PAC DNA was eluted from the Qiagen Tip 500 using 15 ml of buffer QF (1.25 M NaCl, 50 mM Tris, pH 8.5, containing 15% isopropanol) preheated to 65°C. The purified DNA was precipitated by adding 0.7 ml of room temperature isopropanol and centrifuged at $\geq$ 15,000 x g at 4°C for 30 minutes. The supernatant was decanted and the plasmid pellet was allowed to air dry for 5 to 10 minutes. The pellet was resuspended in 200 μl sterile 1x TE. Purified BAC/PAC DNA was electrophoresed on a 1% TBE gel, stained with ethidium bromide, visualized on a transilluminator, and quantified using an Ultrospec 2000 UV spectrophotometer.

Plasmid DNA isolation: I.M.A.G.E. EST clones representing markers mapping to 17p11.2 were obtained (Research Genetics, Incyte Genomics, or the German Genome Project) and grown on LB agar plates containing appropriate antibiotic selection. Individual colonies were then grown in 5 mL LB broth culture containing antibiotic and DNA was isolated from the broth cultures using the Qiagen miniprep kit according to manufacturer's instructions. Briefly, pelleted bacterial was resuspended in 250 μL of buffer P1, lysed with 250 μL of buffer P2 and 350 μL of Qiagen buffer N3 (3 M guanidine-HCl, pH 4.8). Cellular debris was pelleted in a microcentrifuge at maximum speed and the supernatant was removed to a QIAprep spin column, and centrifuged for 30-60 seconds. The column was subsequently washed with 0.5 mL of buffer PB, centrifuged for 30-60 seconds, and buffer PE (10 mM Tris-HCl, pH 7.0, 50% ethanol)

and centrifuged again for 30-60 seconds. The DNA was eluted with 30-50 µL of distilled water or 1x TE buffer. Prepared DNA was run on a 1% TBE gel, stained with ethidium bromide, visualized on a transilluminator, and quantitated using a Ultrospec 2000 UV spectrophotometer.

Southern analysis: Standard Southern analysis was used to map markers to BACs/PACs within the contig and to the hybrid mapping panel. BAC or PAC DNA (6 µg) or human genomic DNA (9 µg) or rodent-human hybrid DNA was digested with 4 U/µg of *EcoRI* for ~16-18 hours at 37°C. The human genomic restriction digest also contained 2.5 mM of spermidine. All digest samples were then electrophoresed in 1% agarose gels in 1x Tris-acetate-EDTA buffer (TAE; 0.04 M Tris-acetate, 1 mM EDTA) with buffer recirculation. Gels were depurinated in two gel volumes of 0.25 N HCl and denatured in two gel volumes of 0.4 N NaOH. DNA was transferred to an Amersham Hybond-N+ nylon membrane in 10x sodium chloride and sodium citrate (SSC; 20x SSC stock solution is composed of 3 M NaCl, 0.3 M Na-citrate) and UV-crosslinked in a Stratalinker 1800 to ensure stability. Filters were prehybridized and hybridized in a solution of 1 M NaCl, 1% SDS, 10% dextran sulfate, and 0.1 mg/mL herring sperm DNA at 65°C or 1% bovine serum albumin or BSA, 1 mM EDTA, 7% SDS, and 0.5 M Na_2HPO_4, pH 7.2 at 50°C. All DNA probes (purified PCR products or plasmid inserts) were labeled with the Amersham Rediprime II kit and unincorporated nucleotides were removed with a spermine and herring sperm DNA precipitation. During preassociation, ~10^6 cpm/mL of probe was annealed to 0.25 mg/mL placental DNA to mask repeat sequences, and then hybridized to the filter for ~16-18 hours. Blots were washed for 20-

30 minutes in 0.1x SSC, 0.1% SDS at room temperature, followed by a stringent wash in preheated 0.1x SSC, 0.1 SDS, at 65°C. The blots were then exposed to X-ray film with one or two intensifying screens at -80°C for 1-5 days.

Database searches: Extensive information about the genes/ESTs in the region was obtained through the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov), including Unigene (http://www.ncbi.nlm.nih.gov/UniGene), the EST database (http://www.ncbi.nlm.nih.gov/dbEST) and the STS database (http://www.ncbi.nlm.nih.gov/dbSTS). Draft assembly of the human genome project data was found at the human genome browser site (http://genome.ucsc.edu) and through Project ENSEMBL (http://www.ensembl.org). Supplementary marker information was derived from the Genome Database (http://www.gdb.org). Sequence alignments and unfinished high-throughput genome sequence BLAST searches were conducted through the Baylor Search Launcher or directly through the NCBI MapViewer. Proteomic analysis was carried out at the Expasy site (http://www.expasy.ch/) and through the Baylor Search Launcher (http://searchlauncher.bcm.tmc.edu:9331/seq-search/protein-search.html). Conserved sequence domains were identified through NCBI (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

PCR: The polymerase chain reaction (PCR) was also used to map markers to the SMS critical interval. A standard PCR reaction was performed in 25 µl volumes. Amplification of ~50-200 ng of template DNA was performed using 1U of *Taq*

polymerase in a cocktail consisting of 0.8 μM of each appropriate primer (Table 7), 0.25 mM dNTPs, and 1x PCR buffer (10x buffer contains 100 mM Tris-HCl pH 8.3, 15 mM MgCl₂, 500 mM KCl, 0.1% gelatin). Amplification was performed in an MJ Research or Applied Biosystems (ABI) thermocycler under the following conditions: initial denature at 94°C for 4 minutes, followed by 30 cycles of 94°C for 1 minute, 55°C for 1 minute, and 72°C for 3 minutes, and a final extension of 72 °C for 10 minutes. PCR products were then electrophoresed on a 1% or 2% TBE agarose gels, stained with ethidium bromide, and visualized on a UV transilluminator.

Northern analysis: The expression pattern of various ESTs was visualized using human multiple tissue, fetal, and brain II northern blots purchased from Clontech and handled according to manufacturer's instructions. Briefly, prehybridizations and hybridizations were performed in 5x SSPE (0.75 M NaCl, 0.05 M NaH₂PO₄, 5 mM EDTA), 10x Denhardt's solution (0.2% BSA, 0.2% Ficoll, 0.2% Polyvinylpyrrolidone), 100 μg/mL herring sperm DNA, 2.0% SDS, 50% deionized formamide at 42°C, with ~10⁶ cpm/mL of each probe. All DNA probes (purified PCR products or plasmid inserts) were labeled with the Amersham Rediprime II kit and unincorporated nucleotides were removed with a spermine and herring sperm DNA precipitation. Filters were hybridized for ~18-20 hours at 42°C and then washed in 2x SSC, 0.05% SDS for ~30 minutes at room temperature with 1-2 changes of solution, then transferred to a wash of 0.1x SSC, 0.1% SDS for ~30-40 minutes at 50°C. The blots were exposed to Amersham Hyperfilm, with one or two intensifying screens at -80°C for 2-5 days.

Chapter III.

RAI1 as a candidate gene for SMS

The physical and transcription map of the SMS critical interval allowed us to identify several promising SMS candidate genes. The second major goal of my research project was to focus on the characterization of one gene, retinoic acid induced 1 or *RAI1*. Our preliminary *RAI1* expression data indicated that this gene was highly expressed in brain as well as many other tissues and may play a role in neuronal development. As the cellular role of this gene is currently unknown, we began a mutation screen of *RAI1* in several patients with an SMS phenotype but no detectable 17p11.2 deletion. We hypothesized that a mutation in a single gene in these patients may produce an SMS phenotype. Our studies subsequently revealed dominant, deleterious *RAI1* mutations in four patients. These findings strongly suggest that haploinsufficiency of the RAI1 protein is sufficient to produce the craniofacial and neurobehavioral features seen in SMS patients.

Rai1/RAI1 cloning and genomic structure

The mouse form of the *Rai1* gene was originally cloned and named *Gt1* by a Japanese group in 1995 (Imai et al. 1995) who were primarily interested in identifying genetic factors affecting neural differentiation. The *Gt1* transcript was found to be significantly up-regulated in mouse embryonal carcinoma cells following retinoic acid (RA) treatment to induce neuronal differentiation (Imai, Suzuki et al. 1995) though it is

not known at which biochemical stage *Gt1* acts within the RA signalling pathway. The mouse *Gt1* cDNA cloned in 1995 is 7222 bp long (GenBank NM_009021), encoding a protein of 1840 amino acids (aa) (Imai, Suzuki et al. 1995). A discussion of the possible role of *Gt1/Rai1* within the RA pathway is provided in Chapter V. Preliminary characterization of *Gt1*, later officially renamed retinoic acid induced 1 or *Rai1*, demonstrated high expression of this transcript in brain as well as several other adult mouse tissues measured on a northern blot, including heart, lung, liver, kidney, thymus, and small intestine. *In situ* hybridization (ISH) localized the *Rai1* mRNA to several regions of the adult mouse brain consisting of neuronal populations and immunohistochemistry (IHC) with an anti-Rai1 antibody identified the Rai1 protein (weakly) in the cytosol and neurites of neurons (Imai, Suzuki et al. 1995). Following this first report identifying *Rai1*, little more was published on this gene until our group demonstrated that the human homolog of *Rai1* (*RAI1*) mapped to PAC RPCI-1 pc253P07 within the central region of the SMS critical interval on human chromosome 17 band p11.2 (Figure 2) (Lucas, Vlangos et al. 2001). As described in Chapter II, we mapped this gene to the SMS region by hybridization of an *RAI1* clone, EST DKFZp434A139 (GenBank AL133649), to the 17p11.2 somatic cell hybrid panel (Figure 3), then subsequently to pc253P07 DNA by human genome project sequence analysis and by Southern hybridization of DKFZp434A139 to *Eco*RI-digested pc253P07 DNA (Figure 4) (Lucas, Vlangos et al. 2001). Further information about the genomic structure of *RAI1* was published in 2001 by Seranski, et al. Using RT-PCR and 5' RACE fragments, this group assembled a cDNA of 5667 bp (GenBank AJ271790) with an 5589 bp open reading frame encoding 1863 aa. According to their analysis, the 5'UTR was not

identified and the coding region of *RAI1* consists of 7 exons over ~ 20kb of genomic sequence (detailed intron-exon sequence was deposited into GenBank under accession AJ271791) (Seranski et al. 2001). The Seranski *RAI1* cDNA sequence and the original mouse *Rai1* transcript share 76% identity. Small regions of sequence similarity were also noted between *RAI1* and the transcriptional coactivator *SPBP* (later renamed *TCF20*) (Rekdal et al. 2000; Seranski, Hoff et al. 2001). At the time of the publication of our transcription map of the SMS critical interval, we could not discern the cellular role of the Rai1/RAI1 protein and no other proteins in the public databases showed significant similarity.

The Seranski group noted that the *RAI1* protein contains polyglutamine and polyserine tracts. As polyglutamine tracts (encoded by the CAG codon) have been found to expand in several neurodegenerative diseases, this group sequenced the CAG repeats from the non-deleted chromosome from SMS patients. Their results demonstrated that the repeats ranged from 12-14 and while slightly different sequence variations existed, the repeats were not found to be expanded or mutated (Seranski, Hoff et al. 2001). The Rouleau research group at the University of British Columbia was also interested in the CAG repeats regions within *RAI1* and the possible role of *RAI1* in schizophrenia (Joober et al. 1999). Joober et al. demonstrated that while no expansions of the CAG repeat within *RAI1* (then called *hGT1*), were identified, the CAG repeat size significantly differed between groups of schizophrenia patients who responded to conventional neuroleptic treatment and those who did not. Various statistical analyses demonstrated that neuroleptic responders had somewhat shorter *RAI1* CAG repeat sizes compared to

controls, possibly indicating that *RAI1* may be involved in the etiology of some forms of

schizophrenia (Joober, Benkelfat et al. 1999). Further analysis of the *RAI1* CAG repeat

was carried out by the same group, who hypothesized that cellular interactions between

the polyglutamine regions of the ataxin-2 and RAI1 proteins may modify the age of onset

of spinocerebellar ataxia type 2 (SCA2) (Hayes et al. 2000). Linear regression analysis

demonstrated a possible effect of *RAI1* on ataxin-2, though no cellular experiments were

done to confirm this hypothesis (Hayes, Turecki et al. 2000). Similar to previous studies,

Hayes and coworkers reported that the number of CAG repeats within *RAI1* ranged from

10-18 and that the repeats were not expanded (Hayes, Turecki et al. 2000).

The genomic structure of *RAI1* presented by Seranski and colleagues did not

agree with many of the ESTs which represent *RAI1* in the NCBI Unigene database

(http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?ORG=Hs&CID=438904)   or   those

aligned in the human genome browser (http://genome.ucsc.edu). It is difficult to assess

whether this is due to a tissue-specific alternatively spliced form of *RAI1*, as the Seranski

paper does not list the tissue library from which their clones were made. The Rouleau

group subsequently published their own genomic structure of *RAI1* in 2003 (Toulouse,

Rochefort et al. 2003), which seems to represent a more common form of the gene, as

many ESTs also have the same intron/exon boundaries. Toulouse and colleagues cloned

the *RAI1* cDNA from SKNSH neuroblastoma cells treated with 10^{-5} M RA and induced

to undergo neuronal differentiation. This group identified 470 bp of 5' UTR, which was

confirmed by the presence of upstream CpG islands and a putative RA response element

within this regulatory region (Toulouse, Rochefort et al. 2003). They redefined the

Figure 7. Human and mouse *RAI1/Rai1* genomic structure.

Spliced *RAI1* transcript variants derived from the latest builds of the human and mouse genome browser (http://genome.ucsc.edu) are shown. The GenBank accession number and the tissue that each cDNA was isolated from is indicated to the right of each genomic structure, though it should be noted that not all cDNA transcripts are full-length and may be based on incomplete EST fragments. GenBank accession CL215311 represents sequence 5' of gene-trapped *Rai1* ES cell line insert from the German Gene Trap Consortium (http://genetrap.gsf.de). CpG islands and a RA-response element were identified in the *RAI1* regulatory region by Toulouse et al.

[1]Toulouse A et al., Genomics 2003. Aug;82(2):162-71.
[2]Seranski P et al., Gene. 2001 May 30;270(1-2):69-76.
[3]Imai et al., Brain Res Mol Brain Res. 1995 Jul;31(1-2):1-9.

Human *RAI1* genome browser (July 2003 build)

Known Genes Based on SWISS-PROT, TrEMBL, mRNA, and RefSeq

RAI1
RAI1
RAI1

RA response element[1]

CpG islands[1]

RAI1
RAI1
RAI1
RAI1

AB058723, brain
AY172136, neuronal cell line[1]
AL834486, testis
AL834468, melanoma
AJ271790, unknown[2]
AL133649, testis
BC021209, rhabdomyosarcoma
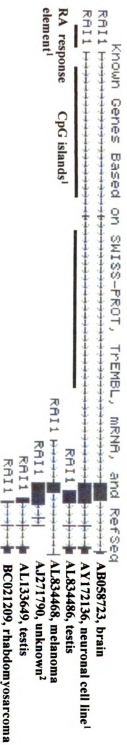
Mouse *Rai1* genome browser (Oct. 2003 build)

AK129449

Rai1

D29801/NM_009021, brain[3]
AK129449, 13 day embryo head
AK013909, embryonic tail
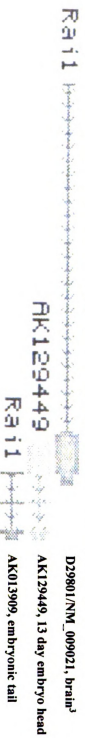
intron/exon junctions that had been proposed by the Seranski group as well as one of the RAI1 isoforms that was present in the NCBI database (XM_016259) and predicted 1452 bp of 3'UTR (Toulouse, Rochefort et al. 2003). The recently proposed RAI1 open reading frame (GenBank AY172136 and GenProt AAO31738) is 5718 bp and encodes an 1906 aa protein (Toulouse, Rochefort et al. 2003). The most common spliced forms of human and mouse RAI1/Rai1 are shown in Figure 7. Based on our own sequence analysis as well as that of Toulouse and coworkers, the major sequence features present in the human RAI1 protein are a nuclear localization signal spanning aa 1160-1240 and a predicted PHD zinc finger domain from aa 1856-1903, similar to a PHD/LAP domain identified in TCF20. If the published mouse Rai1 protein structure (GenProt NP_033047) is correct and the entire coding sequence is present in one exon, the latter motif is unique to the human protein sequence, though there are several mouse ESTs (GenBank AK129449, AK013909) that have 3' intron/exon boundaries which mirror the human RAI1 junctions and would contain the PHD domain (Figure 8). An alignment of the RAI1 aa sequence published by Rouleau (GenProt AAO31738) and the full-length predicted mouse Rai1 protein containing a PHD domain is shown in Figure 8. In addition to the peptides used to generate several antibodies for *in vitro* analysis of RAI1/Rai1, this figure also highlights the putative nuclear localization signal.

Expression pattern of *RAI1/Rai1*

Though there are relatively few publications regarding *RAI1*, there is detailed information documenting the expression pattern of this gene. A review of the published literature as well as our own experiments and information from the public databases

Figure 8. Mouse and human Rai/RAI1 amino acid sequence alignment.

(a) The amino acid sequences of the mouse and human Rai1/RAI1 proteins are aligned using ClustalW and Boxshade programs. The predicted nuclear localization signal is underlined and the putative PHD zinc finger domain is outlined with a box. Peptide sequences used to develop antibodies for immunohistochemistry and western analysis are shown in bold.

(b) Alignment of the putative PHD domain from RAI1 with consensus sequence derived from the NCBI conserved domain database CDD K0G0954. In this alignment, identical amino acids are indicated in red and similar amino acids are in blue. In addition, two putative bipartite NLS sequences from Rai1/Rai1 were identified by similarity to the consensus sequence (KRX$_{10-12}$KRRK).

a.

```
  1   MQSFREERCGFHGKQQNYPQTSQETSRLENYRQPGQAGLSCDRQRLLAKDYYS PQPYTGYE   mouse
  1   MQSFREERCGFHGKQQNYQQTSQETSRLENYRQPSQAGLSCDRQRLLAKDYYN PQPYPSYE   human

 61   GGTGTP SGTVATAAADKYHRGS------KSLQGRPAFP-SYVQDSSPYPGRYSGEEGLQT
 61   GGAGTP SGTAAAVAADKYHRGSKALPTQQGLQGRPAFPGYGVQDSSPYPGRYAGEESLQA

114   WGGPQ P PPPQPQPLPGAVSKYEENLMKKTVVPPPNRQYPEQGPQLPFRTHSLHVP-PPQP
121   WGAPQ P PPPQPQPLPAGVAKYDENLMKKTAVPP-SRQYAEQGAQVPFRTHSLHVQQPPPP

173   QQPLAYPKLQRQKPQNDLASPLPFPQGSHFPQHSQSFPTSSTYAPTVQGGGQGAHSYKSC
180   QQPLAYPKLQRQKLQNDIASPLPFPQGTHFPQHSQSFPTSSTYSSSVQGGGQGAHSYKSC

233   TAPS AQPHDRPMSANANLAPGQRVQNLHAYQPGRLGYEQQQQ----------ALQGRHHT
240   TAPT AQPHDRPLTASSSLAPGQRVQNLHAYQSGRLSYDQQQQQQQQQQQQQQQALQSRHHA

283   QETLHYQNLAKYQHYGQQGQGYCPPDTAVRTPEQYYQTFSPSSSHSPARSVGRSPSYSST
300   QETLHYQNLAKYQHYGQQGQGYCQPDAAVRTPEQYYQTFSPSSSHSPARSVGRSPSYSST

343   PSPLMPNLENFPYSQQPLSTGAFPTGITDHSHFMPLLNPSPTDAASSVDPQAGNCKPLQK
360   PSPLMPNLENFPYSQQPLSTGAFPAGITDHSHFMPLLNPSPTDATSSVDTQAGNCKPLQK

403   EKLPDNLLSEVSLQSLTALTSQVENISNTVQQLLLSKATMPQKKGVKNLVSRTPEQHKSQ
420   DKLPENLLSDLSLQSLTALTLQVENISNTVQQLLLSKAAVPQKKGVKNLVSRTPEQHKSQ

463   HCSPEGSGYSAEPAGTPLSEPPSSTPQSTHAEPQDTDYLSGSEDPLERSFLYCSQARGSP
480   HCSPEGSGYSAEPAGTPLSEPPSSTPQSTHAEPQEADYLSGSEDPLERSFLYCNQARGSP

523   ARVNSNSKAKPESVSTCSVTSPDDMSTKSDDSFQSLHSTLPLDSFSKFVAGERDCPRLLL
540   ARVNSNSKAKPESVSTCSVTSPDDMSTKSDDSFQSLHGSLPLDSFSKFVAGERDCPRLLL

583   SALAQEDLASEILGLQEAIVEKADKAWAEASSLPKDNGKPPFSLENHGACLDTVAKTSWS
600   SALAQEDLASEILGLQEAIGEKADKAWAEAPSLVKDSSKPPFSLENHSACLDSVAKSAWP

643   QPGEPETLPEPLQLDKGGSTKDFSPGLFEDPSVAFATTDPKKTSSPLSFGTKPLLGTATP
600   RPGEPEALPDSLQLDKGGNAKDFSPGLFEDPSVAFATPDPKKTTGPLSFGTKPTLGVPAP

703   DPTTAAFDCFPDTPTASSVDGANPFAWPEENLGDACPRWGLHPGELTKGLEQGAKASDGV
720   DPTTAAFDCFPDTTAASSADSANPFAWPEENLGDACPRWGLHPGELTKGLEQGGKASDGI

763   GKADAHEASACMGFQEDHAIGKPAAALSGDFKQQEAEGVKEEVGGLLQCPEVAKADQWLE
780   SKGDTHEASACLGFQEEDPPGEKVASLPGDFKQEEVGGVKEEAGGLLQCPEVAKADRWLE

823   ESRHCCSSTDFGDLPLLPPPGRKEDLEAEEEYSSLCELLGSPEQRPSLQDPLSPKAPLMC
840   DSRHCCSTADFGDLPLLPPTSRKEDLEAEEEYSSLCELLGSPEQRPGMQDPLSPKAPLIC

883   TKEEAEEALDTKAGWVSPCHLSGEPAVLLGPSVGAQSKVQSWFESSLSHMKPGEEGPEME
900   TKEEVEEVLDSKAGWGSPCHLSGESVILLGPTVGTESKVQSWFESSLSHMKPGEEGPDGE

943   RAPGSSGTSQGSLAPKPNKPAVPEGPIAKKEPVPRGKSLRSRRVHRGLPEAEDSPCRVPA
960   RAPGDSTTSDASLAQKPNKPAVPEAPIAKKEPVPRGKSLRSRRVHRGLPEAEDSPCRAPV

1003  LPKDLLLPESCTGPPQGQAEGAGAPGRGLSEGLPRMCTRSLTALSEPQTPGPPGLTTTPT
1020  LPKDLLLPESCTGPPQGQMEGAGAPGRGASEGLPRMCTRSLTALSEPRTPGPPGLTTTPA
```

Figure 8, continued

```
1063  PPDKLG GKQRAAFKSGKRVGKPSPKAASSPSNPAALPVASDSSPMGSKTKEPDSPSMPGK
1080  PPDKLG GKQRAAFKSGKRVGKPSPKAASSPSNPAALPVASDSSPMGSKTKETDSPSTPGK

1123  DQRSMV LRSRTKPQQVFHAKRRRPSESRIPDCRATKKLPANNHLPTAFKVSSGPQKEGRM
1140  DQRSMI LRSRTKTQEIFHSKRRRPSEGRLPNCRATKKLLDNSHLPATFKVSSSPQKEGRV

1183  SQRVK VPKPGTGNKLSDRPLHTLKRKSAFMAPVPAKKRSLILRSNN-------GSGGDGR
1200  SQRAR VPKPGAGSKLSDRPLHALKRKSAFMAPVPTKKRNLVLRSRSSSSSNASGNGGDGK

1236  EERA ESSPGLLRRMASPQRARPRGSG--EPPPPPPLEPPAACMGLSTQSSLPSAVRTKVL
1260  EERP EGSPTLFKRMSSPKKAKPTKGNGEPATKLPPPETPDACFKLASRAAFQGAMKTKVL

1294  PPRK GRGLKLEAIVQKITSPGLKKLACRVAGAPPGTPRSPALPERR-----PGGSPAGAE
1320  PPRK GRGLKLEAIVQKITSPSLKKFACKAPGASPGNPLSPSLSDKDRGLKGAGGSPVGVE

1349  EGL GGMGQMLPAASG-ADPLCRNPASRSLKGKLLNSKKLSSAADCPKAEAFMSPETLPSL
1380  EGL VNVGTGQKLPTSGADPLCRNPTNRSLKGKLMNSKKLS-STDCFKTEAFTSPEALQPG

1408  GTARAPKKRSRKGRTGTLGPSKGPLEKRPCPGQPLLLAPHDRASSTQGGGEDNSSGGGKK
1439  GTALAPKKRSRKGRAGAHGLSKGPLEKRPYLGPALLLTPRDRASGTQGASEDNSGGGGKK

1468  P KTEELGPASQPPEGRPCQPQTRAQKQPGQASYSSYSKRKRLSRGRGKTAHASPCKGRAT
1499  P KMEELGLASQPPEARPCQPQTRAQKQPGHTNYSSYSKRKRLTRGRAKNTTSSPCKGRAK

1528  RRRQQQVLPLDPAEPEIRLKYISSCKRLRADSRTPAFSPFVRVEKRDAYTTICTVVNSPG
1559  RRRQQQVLPLDPAEPEIRLKYISSCKRLRSDSRTPAFSPFVRVEKRDAFTTICTVVNSPG

1588  DEPKPHWKPSSSAASSSTSS-SSLEPAGASLTTFPGGSVLQQRPSLPLSSTMHLGPVVSK
1619  DAPKPHRKPSSSASSSSSSSSFSLDAAGASLATLPGGSILQPRPSLPLSSTMHLGPVVSK

1647  ALSTSCLVCCLCQNPANFKDLGDLCGPYYPEHCLPKKKPKLKEKARLEGTLEEASLPLER
1679  ALSTSCLVCCLCQNPANFKDLGDLCGPYYPEHCLPKKKPKLKEKVRPEGTCEEASLPLER

1707  TLKGLECSASTTAAAPTTATITTPTALGRLSRPDGPADPAKQGPLRTSARGLSRRLQSCY
1739  TLKGPECAAAATAGKP-------------PRPDGPADPAKQGPLRTSARGLSRRLQSCY

1767  CCDGQGDGGEEVAQADKSRKHECSKEAPTEPGGDTQEHWVHEACAVWTSGVYLVAGKLFG
1785  CCDGREDGGEEAAPADKGRKHECSKEAPAEPGGEAQEHWVHEACAVWTGGVYLVAGKLFG

1827  LQEAMKVAVDMPCTSCHEPGATISCSYKGCIHTYHYPCANDTGCTFIEENFTLKCPKHKR
1845  LQEAMKVAVDMMCSSCQEAGATIGCCHKGCLHTYHYPCASDAGCIFIEENFSLKCPKHKR

1887  LPL
1905  LP-
```

Figure 8, continued

b.

Alignment of putative *Rai1* PHD domain (aa 1856-1903) to consensus PHD from NCBI conserved domain database CDD K0G0954:

```
RAI1  MCSSC E-A AT G C G LHT YP AS  CIFI-------EN SLK PKH
PHD   VCNLC  VKSGAC QCS T RTA VT AF GLEMKTILKEND VK KSY SKH
```

Two putative bipartite NLS (aa 1160-1176 and aa 1223-1239) were identified by BLAST homology searches. The basic or bipartite NLS is composed of two stretches of basic amino acids separated by a spacer of 10-12 amino acids $(KRX_{10-12}KRRK)$[1]:

Rai1 aa 1160-1176: RRRPSEGRLPNCRATKK
Rai1 aa 1223-1239: KRKSAFMAPVPTKKRNL

[1]Robbins J et al., Cell. 1991 Feb 8;64(3):615-23.

demonstrates that Rai1/RAI1 mRNA is expressed on a northern blot in almost all tissues in the adult animal (Lucas, Vlangos et al. 2001; Seranski, Hoff et al. 2001; Toulouse, Rochefort et al. 2003). Following the original cloning of mouse Rai1, Imai et al. demonstrated Rai1 expression in several adult mouse tissues on a northern blot as well as preliminary ISH and IHC results which demonstrate localization to regions of the brain. Our own lab has performed standard northern blots using Clontech adult tissue and fetal blots and identified ~8.0 kb RAI1 mRNA expression in all tissues examined (Figure 6) (Lucas, Vlangos et al. 2001). Seranski and coworkers also utilized Clontech blots to examine RAI1 expression and identified the common ~8.0 kb transcript in all adult and fetal tissues as well as several specific neuronal tissues (cerebellum, cerebral cortex, medulla, bone marrow, occipital pole, frontal lobe, temporal lobe, and putamen). Two other splice variants, a 1.5 kb and a 10 kb transcript, were expressed in most tissues at somewhat lower levels. The Rouleau group performed a similar northern analysis using the Clontech MTN IV brain blot containing amydala, caudate nucleus, corpus collosum, hippocampus, whole brain, substantia nigra, and thalamus and noted that the ~8.0 RAI1 transcript was present in all of these regions of the brain at similar levels except corpus collosum, where no expression was found (Toulouse, Rochefort et al. 2003).

An examination of the current NCBI Unigene database entry for human RAI1 (Hs.438904) confirms that there are at least 135 ESTs representing this gene, isolated from a wide variety of adult and fetal tissues libraries. A large number of these ESTs are from cancer cell lines, suggesting that RAI1 may be highly expressed in several different carcinomas. It is also interesting to find that RAI1 mRNA comprises 1.97% of an

expression library made from adult colon, suggesting the *RAI1* may have a very important regulatory role in the digestive tract (http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Hs&LID=3202). It is possible that RAI1 expression in the colon is very important physiologically, as SMS patients often have gastrointestinal disorders including constipation. The database of Human Unidentified Gene-Encoded Large Proteins (HUGE) is another valuable source of information about *RAI1* expression. On the HUGE website, the Japanese Kazusa Human cDNA Project provides information for entry KIAA1820 (GenBank AB058723; http://www.kazusa.or.jp/huge/gfpage/KIAA1820), which represents one transcript variant of *RAI1* (Figure 7). The expression pattern of KIAA1820 was determined by a combination of RT-PCR cDNA amplification and quantitation by ELISA. The HUGE analysis revealed at least mid-level *RAI1* expression in all tissues examined, high expression in various subsections of the brain (including the corpus collosum, which did not show expression on a northern blot), the kidney, spleen, heart and spinal cord, and very high expression in the whole brain and ovaries. At the current time, it is difficult to determine if RAI1 expression in the reproductive system affects SMS patients, as no patients have been reported to have children and the fertility of these individuals is unknown.

While most of these studies examined the *Rai1/RAI1* mRNA levels, protein expression studies are also underway. Preliminary IHC experiments have been performed by the Rouleau lab, using antibodies generated against the human and mouse Rai1 peptides (Figure 8). Initial studies in P19, Cos-1 and Skbr-3 cells indicate that the Rai1 protein is localized mainly in the nucleus (G. Rouleau, personal communication).

Further western blot and IHC studies in mouse tissues demonstrate that the Rai1 protein is expressed as a ~175 kDa protein in all tissues examined, at especially high levels in neuronal tissues (spinal cord and brain; G. Rouleau, personal communication). Valerie Vinoverski from the Elsea laboratory has also demonstrated positive nuclear localization using an Rai1:GFP fusion protein (data not shown); these preliminary data provide evidence for true nuclear targeting utilizing the nuclear localization signal and possible evidence for the role of Rai1 in transcriptional regulation. Further experiments such as mutation of the putative NLS in the Rai1:GFP fusion will provide further evidence that Rai1 is truly targeted to the nucleus.

The developmental expression pattern of *Rai1/RAI1* is less well-characterized. As stated above, expression on human fetal northern blot (Figure 6) and various ESTs suggest that *RAI1* is expressed during embryonic stages, but a full developmental profile has not yet been established. The mouse Unigene entry (Mm.275497) lists fewer ESTs that than the human entry, many derived from whole brain, though some have been isolated from mouse embryo (GenBank AK013909, BQ832285, AI893920), indicating that *Rai1* may be embryonically expressed. We have purchased a mouse *Rai1* 3' end cDNA clone (IMAGE:1211624) from heart and used this radiolabeled cDNA to probe a mouse full-stage conceptus northern blot (representing embryonic days E4.5-E18.5). As shown in Figure 9, an ~8.0 kb *Rai1* transcript is evident at all mouse developmental stages, with highest expression from embryonic days E9.5-E15.5. A more definitive developmental profile of the Rai1/RAI1 mRNA and protein expression pattern is underway using ISH and IHC and will be necessary for determining whether Rai1/RAI1

Figure 9. Mouse _Rai1_ embryonic expression pattern.

The EST IMAGE: 1211624, representing the middle and 3' end of mouse _RAI1_, was hybridized to a SeeGene full-stage embryonic northern blot. An ~8.0 kb transcript is evident at all time points and is strongest from E10.5-E15.5. _Gapdh_ is included as a mRNA loading control.

Rai1

~8.5 kb

E4.5
E5.5
E6.5
E7.5
E8.5
E9.5
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
E17.5
E18.5

Gapdh

1.0 kb
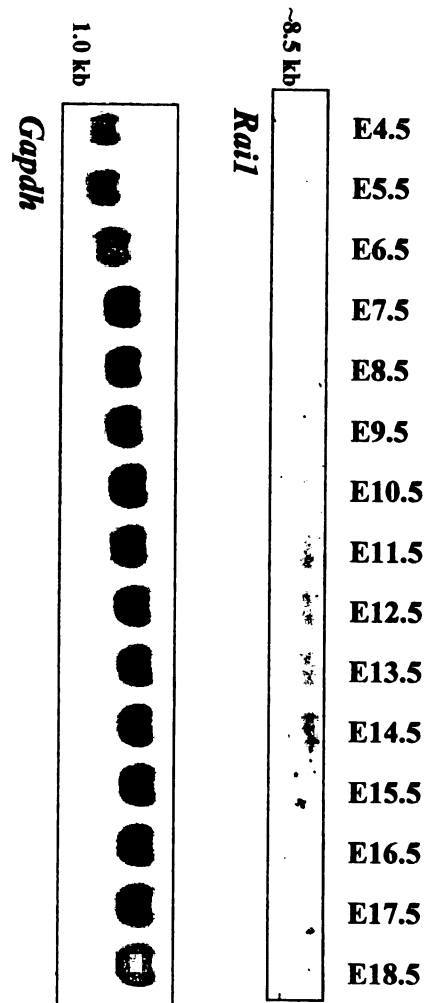
has a global impact on early development and whether haploinsufficiency of Rail/RAI1 could cause neuronal as well as craniofacial defects. *RAI1* may be especially important to the development of higher organisms, as putative *RAI1* EST homologs have been sequenced from rat (GenBank BU758552), *Bos taurus* (GenBank BE667697), *Sus scrofa* (BQ598394), *Gallus gallus* (BU138127), and *Xenopus laevis* (GenBank BG813716); to date, no orthologs have been identified in *Drosophila, C. elegans* or yeast.

Clinical description of putative SMS patients with no cytogenetic deletion

As the true cellular function of *RAI1* is unknown and not currently measurable, our laboratory devised another method to assess *RAI1* as a potential candidate gene. We hypothesized that patients with features consistent with the SMS phenotype but in whom no 17p11.2 deletion could be detected may harbor a mutation in a single candidate gene which may be primarily responsible for their phenotype. As described in the Chapter I, there is precedent for this type of analysis in several microdeletion syndromes, such AGS in which haploinsufficiency of a single gene (*JAG1*) due to mutation or deletion has been found to be primarily responsible for the AGS phenotype. We began our mutation screening by identifying an initial pool of 12 patients with many features consistent with SMS but who did not harbor a 17p11.2 deletion detectable by routine cytogenetics or FISH analysis (Tables 5, 6).

As shown in shown in Tables 5 and 6, these putative SMS patients with no cytogenetic deletion display many of the typical features of SMS such as craniofacial abnormalities, flat midface, sleep disturbances, and characteristic behavioral features such as self-abusive and explosive episodes, as well as hyperactivity. However, these

patients do not display many of the more variable characteristics seen in <50% of typical SMS patients, such as gross organ abnormalities including heart and renal defects. SMS 129, SMS156, SMS159 (Table 5) were part of an initial evaluation of individuals placed at the Elwyn Institute residential care facility based on a strong clinical suspicion of SMS. A detailed clinical summary of SMS129, SMS156, and SMS159 contributed by Brenda Fincucane, a genetic counselor who evaluated these patients, is described within Materials and Methods. SMS188 was referred to Dr. Elsea by a clinical from Belgium (Koenraad Devriendt), who suspected this patient had SMS because of his explosive and self-abusive behavior and developmental delay (Table 5). However, some of the patients in this preliminary pool (Table 6), such as SMS117, who has widely spaced fingers and toes, and SMS119, who was exposed to drugs and alcohol in utero, display inconsistent features and may not truly have SMS.

FISH experiments

All putative SMS nondeletion patients in our mutation screen were evaluated by Christopher Vlangos in the Elsea lab for sub-microscopic deletions within 17p11.2 using an extensive series of FISH probes that span the SMS critical region (Figure 10 and data not shown). In Figure 10 panel (a), an abbreviated contig of the SMS critical interval demonstrates several of the BACs, PACs, and cosmids used for our FISH experiments. Representative FISH results using genomic clones from within the SMS critical interval are shown for SMS 129 in Figure 10 using BAC probe CITC-40I23 (representing the PEMT2 gene) in panel (b) and cosmid probe c62F2, which contains the FLII gene, in panel (c). While these FISH analyses do not completely eliminate the possibility of a

Figure 10. FISH analysis on putative SMS patients with no detectable deletion.

(a) An abbreviated contig of the SMS critical interval (as of 2003) is shown, highlighting *RAI1* in blue. Several BACs used for FISH to evaluate SMS deletions are shown beneath the contig, as well as various overlapping cosmids (dashed lines) also used for FISH probes. BAC CITC-40I23 (containing *PEMT2*) and cosmid 62F2 (containing the *FLII* gene) are shown in red and were used as probes in the FISH experiments shown in panels b and c.

(b) FISH analysis of SMS129 showing no deletion of the green test probe (bc40I23, indicated by white arrows) performed by Christopher Vlangos of the Elsea laboratory. The red control probe (RPCI11-314M5) was used to identify the q arm of chromosome 17.

(c) FISH analysis of SMS129 showing no deletion of the green test probe cosmid c62F2 (from the chromosome 17 specific cosmid library) performed by Christopher Vlangos of the Elsea laboratory. The red control probe (RPCI11-314M5) was used to identify the q arm of chromosome 17. Similar experiments were performed on all of the nondeletion patients described in Tables 5 and 6 to determine that these individuals did not harbor 17p11.2 deletions.

a.

17p11.2

tel

CTC bc40123
RPCI-11 bc915D19
RPCI-1 pc253P07
RPCI-11 bc384M20
RPCI-11 bc189D22
RPCI-1 pc178F10

PEMT2
SMCR2
SMCR3
SMCR4
SMCR5
RAI1
SREBF1
TOM1L2
DKFZp586M1120
ATPAF2
MGC 3048
DRG2
MYO15A
FLJ20308
LLGL1
FLI1
SMCR7
TOP3A
SMCR8
SHMT1

cen

b.

c.
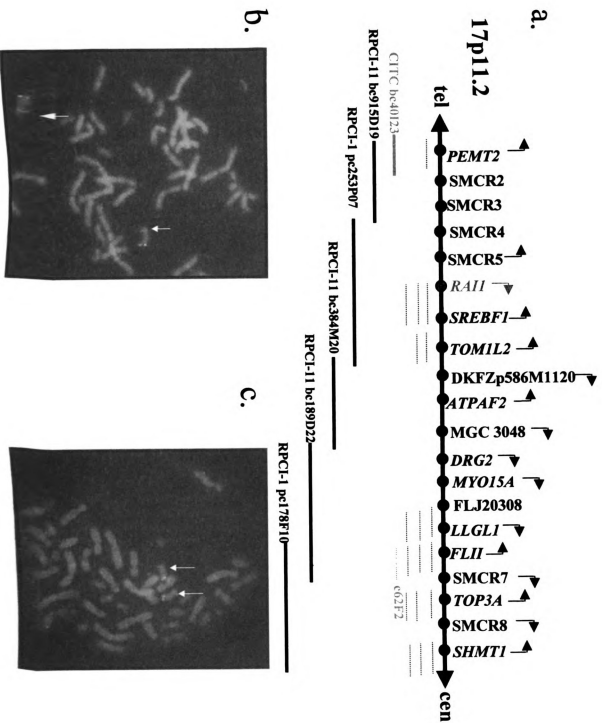
Table 5. Phenotypic characteristics of SMS patients without a cytogenetic deletion and putative SMS patients with deleterious RA11 mutations.

Features of an SMS patient with a typical 17p11.2 deletion (SMS126) and a small 17p11.2 deletion (HOU142-540) are contrasted to putative SMS patients with no cytogenetic deletion who harbor deleterious RA11 mutations (SMS129, SMS156, SMS159, and SMS188). Phenotypic information was derived from www.genetests.org; + = positive, - = negative, N= unknown or not evaluated. In the interest of space, the gender and age(s) of evaluation of each patient are listed below:

a Female, age of evaluation 41 years
b Male, ages of evaluation 10, 14, 15 years
c Male, ages of evaluation 14, 30 years
d Female, ages of evaluation 17, 21, 31 years
e Male, ages of evaluation 13, 19 years
f Male, age of evaluation 12 years

Features	SMS126[a]	HOU142-540[b]	SMS129[c]	SMS156[d]	SMS159[e]	SMS188[f]
Features common in >75% of SMS Patients						
Craniofacial/Skeletal						
Brachycephaly	·	+	+	+	+	+
Midface hypoplasia	+	+	+	+	+	+
Prognathism (relative to age)	+	+	+	+	+	+
Tented upper lip	+	+	+	+	+	+
Broad, square face	+	+	+	+	+	+
Brachydactyly	+	·	·	·	·	+
Short stature	+	·	+	+	+	+
Chronic ear infections	+	+	·	+	+	+
Hoarse, deep voice	+	+	+	+	+	+
Neurological/Behavioral						
Mental retardation	+	+	+	·	+	+
Speech delay	+	·	·	·	·	+
Motor delay	+	·	·	·	+	+
Infantile hypotonia	+	·	·	·	+	N
Sleep disturbance	+	+	+	+	+	N
Self-hugging/hand wringing	+	+	+	+	+	+
Onychotillomania	+	+	+	+	+	+
Polyembolokoilamania	+	+	+	+	+	+
Head banging / face slapping	+	+	+	·	·	+
Hand biting	+	+	·	·	+	+
Attention seeking	+	+	+	+	+	+
Features common in 50 – 75% of SMS patients						
Hearing loss	+	+	+	·	·	+
Ocular abnormalities	+	+	·	+	+	+
Synophrys	+	+	+	+	+	+
Scoliosis	+	+	+	·	+	N
Features common in < 50% of SMS patients						
Cardiovascular anomalies	+	+	·	+	·	·
Renal anomalies	·	+	·	·	·	·
Seizures	*	+	*	·	·	·
Cleft lip/palate	·	·	·	·	·	·

Table 6. Phenotypic characteristics of putative SMS patients with no cytogenetic deletion who harbor no obvious *RAI1* mutation.

Phenotypic information was derived from www.genetests.org; + = positive, - = negative, N= unknown or not evaluated. In the interest of space, the gender, any phenotypic features not consistent with SMS, and the age of evaluation of each patient is presented below:

[a]Female with proximally placed fingers and widely spaced toes, age of evaluation 16 years

[b]Female with drug and alcohol exposure *in utero*, age of evaluation 6.5 years

[c]Female, age of evaluation 17 years

[d]Female, age of evaluation 17 years

[e]Female, age of evaluation not known

[f]Female, age of evaluation 21 years

[g]Male, age of evaluation 6 years

[h]Female, age of evaluation 10 years

Features	SMS 117[a]	SMS119[b]	SMS 135[c]	SMS153[d]	SMS172[e]	SMS196[f]	SMS198[g]	SMS199[h]
Features common in >75% of SMS Patients								
Craniofacial/Skeletal								
Brachycephaly	+	+	+	+	+	+	+	+
Midface hypoplasia	+	+	+	+	+	+	+	+
Prognathism (relative to age)	+	·	+	+	N	+	+	+
Tented upper lip	·	+	+	+	+	+	+	+
Broad, square face	+	·	·	·	+	·	+	+
Brachydactyly	+	·	·	+	+	+	+	+
Short stature	+	+	+	+	+	+	·	+
Middle ear/laryngeal anomalies	N	+	+	N	N	N	+	+
Hoarse, deep voice	+	·	+	+	N	+	+	+
Neurological/Behavioral								
Mental retardation	+	+	+	+	+	+	+	+
Speech delay	+	+	+	+	+	+	+	+
Motor delay	+	+	+	+	+	+	+	+
Infantile hypotonia	+	+	+	+	N	N	+	+
Sleep disturbance	+	·	+	+	+	+	+	+
Self-hugging/hand wringing	N	+	+	+	+	+	+	+
Onychotillomania		·	·	·	·	·	·	·
Polyembolokoilomania	·	+	+	+	+	+	+	·
Head banging / face slapping	N	+	+	+	N	N	N	+
Hand biting	N	+	N	+	+	+	N	+
Features common in 50 – 75% of SMS Patients								
Attention seeking	N	+	+	+	+	+	+	+
Hearing loss	+	N	+	·	N	N	N	N
Ocular abnormalities	+	+	+	+	+	+	+	+
Synophrys	+	·	·	N	·	N	N	·
Scoliosis	+	·	·	+	N	N	N	+
Features common in < 50% of SMS patients								
Cardiovascular anomalies	·	·	·	·	·	·	N	+
Renal anomalies	+	+	+	·	N	N	N	+
Seizures	*	+	+	·	·	·	·	·
Cleft lip/palate	·	·	·	·	·	·	·	·

85

small, cryptic deletion within 17p11.2, they demonstrate that the *individuals screened for* single gene mutations did not harbor detectable deletions within the *SMS critical region.*

Sequencing of *RAI1* in putative SMS patients with no deletion

Having identified 12 putative SMS patients without detectable deletions (Tables 5-6), we undertook systematic PCR and sequencing of genes within the SMS critical region using DNA from our non-deletion patient pool, beginning with three genes first localized to the SMS critical interval in 1997, *DRG2*, *RASD1*, and *RAI1*. *RAI1* primer sequences spanning the entire genes are shown in Table 7. While our initial sequencing identified several known SNPs, no deleterious mutations were found in either *RASD1* or *DRG2* in any of our patients. However, a 29 bp deletion within exon 3 of *RAI1* was identified on one allele from SMS129 (Figure 11). This deletion was clearly evident on a 2% agarose gel following PCR amplification with exon 3 primers, as two bands representing the deleted and non-deleted alleles were resolved on the gel (Figure 11, panel b). We isolated and purified each band from the gel and sequenced them separately. As shown in the *RAI1* mRNA coding sequence in Figure 11, panel c., the 29 bp deletion produces a frameshift that introduces 8 incorrect amino acids, followed by a premature stop codon, truncating the protein. This mutation also abolishes a *Psp*OMI restriction site (Figure 11, panel b). We predict that this deleted allele encodes a truncated and either abnormally-functioning or non-functional RAI1 protein, likely resulting in haploinsufficiency for RAI1. We sequenced the exon 3 amplimer from the parents of SMS129, SMS127 and SMS129 as well as one sibling, SMS164 and subjected this amplimer to restriction digest with *Psp*OMI. As shown in Figure 11, panel b., the 29

Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temp. (°C)
CCTTCCTCCCTCCCTCCCTTCC	CACCCCTGCAGGTAGTGGCTG	65
CGCTATGCTGGTGAGGAGAGCC	CCGACTGGTAGGCATGAAGATTC	64
CCATGACGAGGCGCGAGACTGC	CAGGGAGGCTGTGTCCTTGGAAG	62
CTGACCACAGCCACTTCATGCC	**CACGGACTCGGGCCTTGGCCTTCG**	63
CAGCTTCTCTTACTCCAACCAG	GCGAAGGCCACCGGAAGGGTCTTC	60
GCCCGACTCCTTTGACACCATCAG	CCGGTCAGCCCTGGGCCACTCGG	65
GGACTTCAAGCAGGAGGAGGTCG	**CACGAGGCGTCCGCAGGTGTG**	64
CACATGAAGCAGTTGAAGAGG	CTGGAGGCAGCCTTGGGTGAC	65
CGTTCTCTCACCGGCCCTGAGTG	GCCACTGGCGTTGCTGCTGCTGC	68
GCGCTCAAAGGAAGTGCGGCC	**CCACATTTACCAGGCCTTCTTCC**	64
CCCTTCCGACAAAAGCACGTCG	GTGTGGCCTCGGCTGTTTCTGTG	64
TGGACTCTCCAAAGGCCCGCT	AGGCCCCAAGTGCATCGTGG	60
CCTGCCGCCACACTCCCTTGGAGG	**CTGCCGGAGCCTCCTTGTTGCAC**	64
TGTGCAGCGCGCGCCACT	ACTCTGCAGATTGTCCGGAGA	57
GCAACACCACCACCCTCACT	AATGCCTCATTTCCATGTCC	62
GCTTGAGGGCTGGGCTCCAAC	CAAAGGCCCAACCTCCAATACC	64
GGACTGTCGAAGGAGGTGCGAGG	GGACTGGAGTGGAGTGTGGAGG	66
GAGGCTCCTGTGCTACTTTGCC	GTTGACACAGCCCAACCATGTGC	64
GCACATGGTTGGGCTGTGTCAAC	GTCAATAAAGATACAACGATTG	62
CAGCTCGATACACACAAATCTTC	CGTTGTTGCCACCACCAGGGACC	64
GGTCCCTGGTTGGTGCACAACCG	GTGGGAGACGGCTTTGTCCTGG	64
CAGGACAAAGCGCGTCTCCCAC	GACTGTGAAGTCCGGAGGTCGTC	57
GAGTCCTCTGGCACCGAACGAG	GCCGCCTCTCGGCAGCCACTCTG	62
CTTGCAGCCCCGGACTCC	TTGCAAGCGGTGGCGAGGAG	62
CCCACCACCACAAAGCA	GGGCTCTTGCTCTCCTTCGT	59
CAAATGTCACCCTCGCGTCC	GACCTGGGAGAGCTCTGTAG	62
TGCTAGGCTGGTGGTGGAAAGG	CGGGATCTAGAAACTGGAAAGG	62

Table 7. *RdII* primers used this this study. *RdII* primers were obtained from the Schmidt lab at NCI-Frederick or designed by the Elsea lab. These primers cover primary transcripts as well as putative splice variants. The annealing temperature for the primers is indicated and Qiagen Q solution or Invitrogen PCR enhancer was routinely added to the PCR mix to amplify difficult templates. **Amplimers in bold have identified mutations.**

Figure 11. SMS129 *RAI1* mutation analysis.

(a) SMS129 pictured at age 30.

(b) Pedigree analysis of the SMS129, his parents (SMS127 and SMS128) and one sibling (SMS164). The RA13/14 PCR product is present in lanes 1, 3, 5, and 7; note doublet in affected individual SMS129. The *Psp*OMI digestion of the RA13/14 PCR product is shown in lanes 2, 4, 6, and 8; note undigested mutant allele in affected individual SMS129 which is not present in the parents or sibling. SMS128 was not noted to carry the 29 bp deletion by PCR, restriction digest, or direct sequencing, though further DHPLC analysis showed ~20% mosaicism in this parental sample (see Figure 16). The 29 bp mutation was not detected in 102 Caucasian control samples which were analyzed by PCR amplification and *Psp*OMI restriction digestion.

(c) The sequence tracing represents the the 29 bp deletion detected following direct sequencing of the RA13/14 amplimer of *RAI1* from SMS129, as well as the wild type sequence detected on the other allele; this deletion eliminates a *Psp*OMI restriction site, misincorporates 8 amino acids (4 of which are shown in the diagram) and produces a downstream stop codon. Other mutation screening results show that SMS129 is heterozygous (C/T) for known SNP in exon 4 (dbSNP:rs3818717).

a.

SMS129

age 30

b.

493 bp →

SMS128   SMS127   SMS129   SMS164

1 2 3 4 5 6 7 8

132 bp

361 bp

c.

5'-AGGGGAGTCC|...|CGAGTCAAA -3'

G   E   S   V   I   L   L   G   P   T   V   G   T   E   S   K : *Normal*

5'-AGGGGAGTCCGCGAGTCAAAGG -3' : *Mutant*

G   E   S   R   V   K

d SMS128) and one
1, 3, 5, and 7; note
f the RA13/14 PCR
nt allele in affected
MS128 was not noted
cing, though further
ee Figure 16). The
hich were analyzed

following direct
s the wild *type*
restriction site,
nd produces a
t *SMS129* is

89

bp deletion was not evident in these individuals, though we could not rule out low-level mosaicism by PCR and sequencing methods. We did pursue further analysis of the exon 3 amplimer from SMS128 by Transgenomic DHPLC mutation screening, which revealed that the 29 bp mutation was present at ~20% level within the genome of this phenotypically normal parent (Figure 16). We also screened >200 Caucasian control chromosomes by PCR and PspOMI restriction digest and did not detect this particular mutation (data not shown).

We then examined three additional patients, SMS156, SMS159, and SMS188 for mutations in RAI1. In SMS156, we identified a deletion of a single cytosine on one allele within exon 3 (Figure 12, panel c), which occurred in a run of 6 C's ending at nucleotide position 5265 of the RAI1 mRNA. This 5265delC on the coding strand produces a subsequent frameshift, introducing 74 incorrect amino acids, abolishing a BglI recognition site, and truncating the protein (Figure 12, panel b). As in the case of the 35delG mutation in connexin 26 which is implicated in recessive hearing loss (Zelante et al. 1997), we were unable to determine absolutely which C is deleted within this run of 6 C's in RAI1. As shown in Figure 12, panel b., the parents of SMS156 (SMS154 and SMS 155) as well as >200 Caucasian control chromosomes screened by PCR and BglI restriction digest (data not shown) did not carry the 5265delC mutation. Again, we predict that this frameshift mutation produces an abnormal or truncated RAI1 protein, which ultimately creates haploinsufficiency of this crucial developmental molecule.

Figure 12. SMS156 *RAI1* **mutation analysis.**

(a) SMS156 pictured at age 31.

(b) The sequence tracing from the RA25/26 *RAI1* amplimer reveals the 5265 delC in exon 3 on one allele; this mutation eliminates a *BgII* restriction site, misincorporates 74 amino acids, eliminates a *BgII* restriction site, and produces a downstream stop codon. Other sequencing results for SMS156 reveal that this individual is heterozygous (C/T) for known SNP in exon 4 C/T (dbSNP:rs3818717) and homozygous (G to C) for known SNP within a non-coding region (dbSNP:rs2297508).

(c) Pedigree analysis of the SMS156, her parents (SMS154 and SMS155). The RA25/26 PCR product is present in lanes 1, 3, and 5 and the *BgII* digestion of the RA25/26 PCR product is shown in lanes 2, 4, 6, and 8 (note undigested mutant band present in SMS156 which is not present in parental samples). The 5265delC mutation was not detected in 110 Caucasian control DNA samples which were analyzed by PCR amplification and *BgII* restriction digestion.
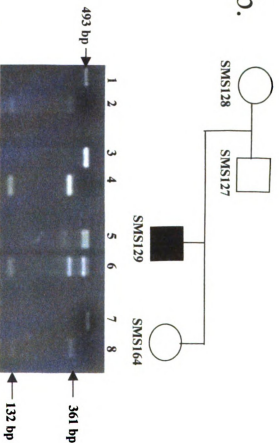
a.

SMS156

age 31

b.


```
SMS154 ○────┐
              │
SMS155 □──────
              │
SMS156 ● : Normal
```

496 bp

1 2 3 4 5 6

171 bp
325 bp

c.
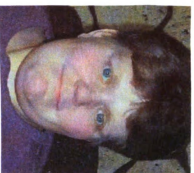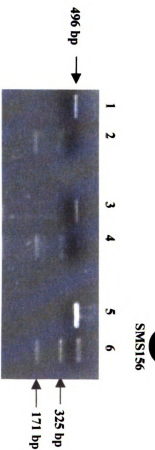
5'-GGGAAGCCCCC

```
  G  K  P      P  R  P  D  G  P  A  D  P  A  K  Q  G  : Normal
CAGGCCTGACGGCCCAGCCTGACCCGGCCAAGCAGGGC -3'
AGGCCTGACGGCCCAGCTGACCCGGCCAAGCAGGGCC -3'
  P  G  L  T  A  Q  L  T  R  P  S  R  A  : Mutant
```
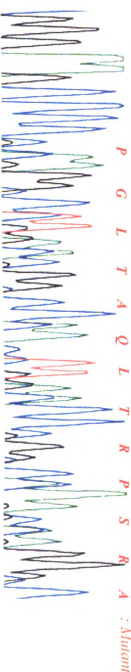
5265delC

als the 5265 delC in
e, misincorporates 74
nstream stop codon.
terozygous (C/T) for
o C) for known SNP

55). The RA25/26
the RA25/26 PCR
resent in SMS156
s not detected in
amplification and

Figure 13. SMS159 *RAI1* mutation analysis.

(a) SMS159 pictured at ages 11 and 19.

(b) The 1449delC mutation in *RAI1* exon 3 on one allele is shown, which misincorporates 34 (12 of which are shown in the sequence tracing) amino acids and produces a premature stop codon. As no known restriction site was altered by the *1449delC* mutation, amplified and sequenced the amplimer containing the *1449delC mutation from* ≥100 Caucasian samples (including the parents of SMS159), and this mutation was not detected in this population.

a.

SMS159

age 11 age 19

b.

S Q H C S P E G S G Y S A E P A G
5'-GCCAGCACTGCAGCCC CGAGGGAGCGGCTACTCAGCCGAGCCCGCAGGC-3' : *Normal*

P R G A A T Q P S P Q A
GAGGGAGCGGGTACTCACCCGAGCCCGCAGGCA-3' : *Mutant*
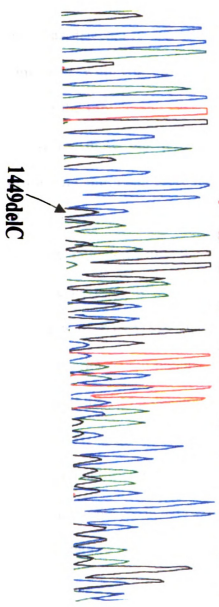
1449delC

h misincorporates
and produces a
y the 1449delC
C mutation from
utation was not

Similarly, a deletion of a single cytosine within a run of 4 C's beginning at nt position 1449 of the *RAI1* mRNA sequence (Figure 13) was found on one allele in patient SMS159 and a deletion of a single cytosine within a run of 5 C's beginning at nt position 3801 was detected in one allele of SMS188 (Figure 14). These deletions also produce frameshifts within the *RAI1* coding sequence, and premature stop codons which can truncate the protein and cause cellular haploinsufficiency. As no restriction sites were altered by these deletions, we directly sequenced 100-200 PCR amplimers from Caucasian control chromosomes as well as parental samples from the family of SMS159, and the mother and two sisters of SMS188. We did not detect the 1449delC or the 3801delC mutation within this population (data not shown). The deleterious mutations in *RAI1* we have identified affect all known putative transcripts from this gene with the exception of KIAA1820 (GenBank AB058723), as the protein is truncated in each mutation prior to the regions of alternative splicing (Figures 7 and 15, panel b). The location of the 29 bp deletion, 1449delC, 3801delC, and 5265delC mutations, as well as several *RAI1* single nucleotide polymorphisms (SNPs) mapped in reference to the *RAI1* sequence published by Toulouse et al. (GenBank AY172136) is shown in Figure 15. A schematic of the predicted RAI1 protein truncations which may arise from the deleterious *RAI1* mutations is also shown in Figure 15, panel b.

Other *RAI1* sequence features

While no other obviously deleterious *RAI1* mutations were identified in our sequencing population, several presumably benign *RAI1* single nucleotide SNPs were identified (Table 8 and Figure 15), which occur in putative SMS patients as well as

Figure 14. SMS188 *RAI1* mutation analysis.

The 3801delC mutation in *RAI1* exon 3 on one allele is shown, which misincorporates 46 (10 of which are shown in the sequence tracing) amino acids and produces a premature stop codon. As no known restriction site was altered by the 3801delC mutation, we PCR-amplified and sequenced the amplimer containing the 3801delC mutation from >75 Caucasian samples (including the mother and two sisters of *SMS188*), and this mutation was not detected in this population.

5'-GAGGGTTCCC

CACCCTCTTCAAGAGGATGTCTTCTCCCAA -3'

P   T   L   F   K   R   M   S   S   P   :Normal

3801delC

ACCCTCTTCAAGAGGGATGTCTTCTCCCAAG -3'

P   P   S   S   R   G   C   L   L   P   :Mutant

normal control chromosomes. As shown in Table 8, known SNPs *(those found in the*

NCBI SNP database: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp) include a

synonymous change in exon 4 (dbSNP:rs3818717) present in several individuals and a

SNP within a non-coding region of the *RAI1* 3' UTR (dbSNP:rs2297508) (Figure 15).

Our sequencing results have also revealed two unclassified SNPs, including a

synonymous change in exon 3 identified in DNA from patient SMS196 and a

nonsynonymous change in exon 3, which produces an R to G amino acid change (Table

8). *The latter* sequence change was isolated by Transgenomic, Inc. denaturing high-

performance liquid chromatography (DHPLC) analysis of control and parental

chromosomes (Figure 16). As this normal individual is heterozygous for the SNP, this

amino acid change is presumably benign, though the true cellular effect of this sequence

change (or any missense mutation) on the RAI1 protein is difficult to assess. We also

analyzed the structure and number of *RAI1* CAG repeats within our patient and control

population. Our studies demonstrated that the number of CAG repeats ranged from 10-

15, which correlates with other published *RAI1* information (Joober, Benkelfat et al.

1999; Hayes, Turecki et al. 2000; Seranski, Hoff et al. 2001). We commonly found

different numbers of CAG repeats on each allele from a single individual (data not

shown), though we did not identify expanded repeats.

One other interesting sequencing feature involving a possibly polymorphic amino

acid repeat was discovered in a control sample. Following the detection of the 3801delC

mutation from an *RAI1* exon 3 PCR product amplified from SMS188 DNA, we

sequenced the same amplimer in >100 control chromosomes. Although none of these

Patient number	Deleterious *RAI1* mutation	Synonymous SNP within coding sequence	Nonsynonymous SNP within coding sequence	SNP within noncoding sequence
SMS117		dbSNP:rs3818717 (T to C)		
SMS119		dbSNP:rs3818717 (T to C)		
SMS129	29 bp del	dbSNP:rs3818717 (T/C het.)		
SMS133		dbSNP:rs3818717 (T to C)		
SMS153		dbSNP:rs3818717 (T to C)		
SMS156	3265delC	dbSNP:rs3818717 (T to C)		
SMS159	14490delC			dbSNP:rs2297508 (G to C)
SMS172				
SMS188	3801deliC			
SMS196		exon 3 (T/C het)		
SMS198				
SMS199		dbSNP:rs3818717 (T/C het)		
Controls		dbSNP:rs3818717 (T to C)	exon 3 (G/C het.; R to G aa change)	dbSNP:rs2297508 (G/C het)

Table 8. *RAI1* sequence changes detected within patient pool of putative SMS patients with no cytogenetic and normal controls.

Figure 15. *RAI1* sequence changes.

(a) The most recently described *RAI1* genomic structure is shown (Toulouse et al., 2003) in this schematic derived from the genome browser (http://genome.ucsc.edu) and mutations identified in our patient population are numbered in agreement with the nucleotide positions of this new genomic structure, beginning with the translation start codon of GenBank accession AY172136 as nt 1. Single nucleotide polymorphisms are also indicated and described in the table below the genomic structure.

(b) A schematic is shown of the normal human RAI1 protein (N terminal at the far left and C terminal at the far right) and predicted truncated RAI1 protein structures which may arise from the dominant, frameshift *RAI1* mutations identified in SMS129, SMS159, SMS156, and SMS188. Colored bars indicate motifs within the RAI1 protein: blue represents the polyglutamine tract, green represents the nuclear localization signal, yellow represent the polyserine tract, purple represents the PHD domain, and red represents incorrect aa which are incorporated downstream of the frameshift mutations. All mutations are predicted to eliminate the PHD domain.

a.

E1 E2 29 bp del 3801delC SNP A SNP C E3 E4 E5 E6 SNP B SNP D 1449delC 5265delC

SNP A	unclassified SNP; nonsynonymous change in exon 3 identified in a control individual (G/C; R to G aa change)
SNP B	unclassified SNP; synonymous change in exon 3 (T/C)
SNP C	dbSNP:rs3818717; synonymous change in exon 4 (T/C)
SNP D	dbSNP:rs2297508 occurs in 3' non-coding sequence (G/C)

b.

Poly-Q NLS Poly-S PHD

Normal human RAI1 protein

SMS159

SMS129

SMS188

SMS156

101

Figure 16. DHPLC *RAI1* mutation screening results from parental and control samples.

Analysis by Transgenomic DHPLC of parental and randomly selected Caucasian control samples detected ~20% mosaicism for the 29 bp deletion within the SMS128 parental sample (mother of SMS129) in the exon 3 amplimer, as well as an unclassified SNP in a normal Caucasian control (G/C heterozygote which results in a nonsynonymous R to G aa change) in the same amplimer.

Fluorescence (mV)

Time (minutes)

Legend:
— Exon 3 SMS128 (parent of SMS129) sample mosaic for 29 bp del
— Exon 3 control sample
— Exon 3 control with SNP present
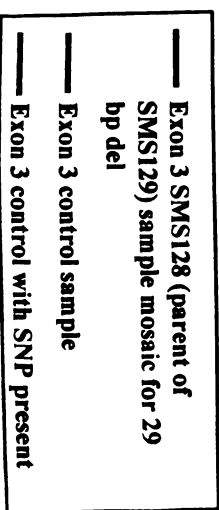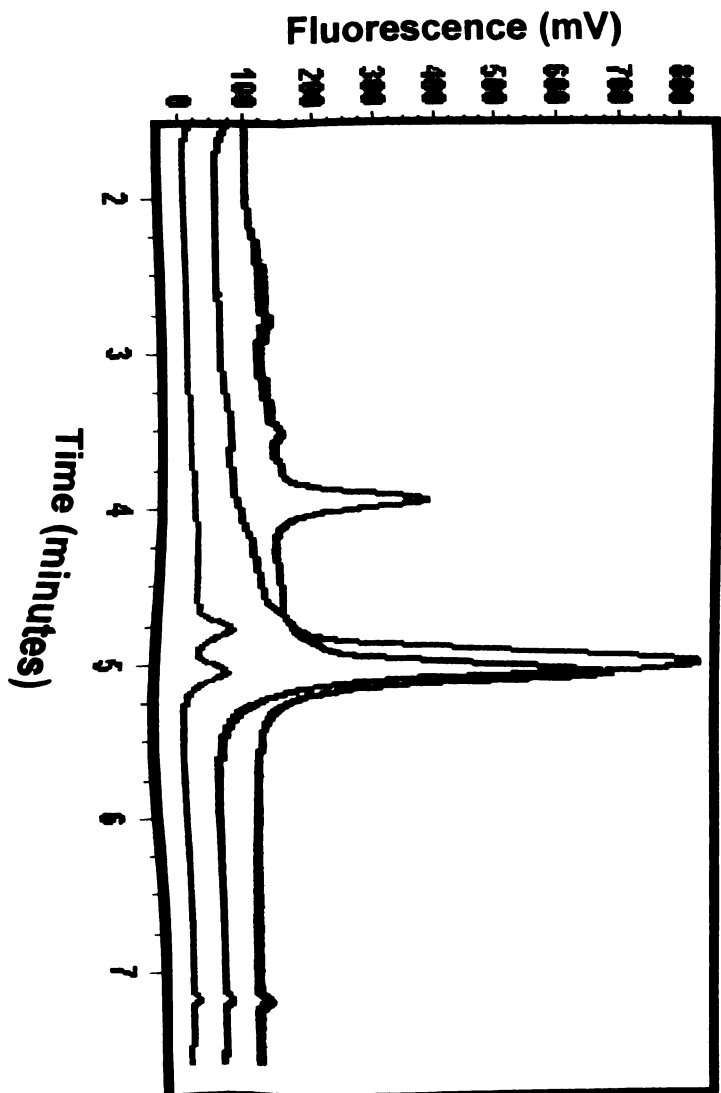
controls contained the 1449delC deletion, one sample did contain a unique polymorphism: one allele coded for two adjacent glutamic acid moieties at aa position 1260 and 1261 and one allele coded for only one glutamic acid (data not shown). The effect of this change on the cellular role of the RAI1 protein, as well as any of the identified SNPs, is unknown at this time, and awaits the result of future functional studies.

Towards high-throughput *RAI1* mutation screening

In order to facilitate more efficient, high-throughput *RAI1* mutation screening, we performed preliminary DHPLC and temperature gradient capillary electrophoresis (TGCE) experiments to determine whether these techniques could identify known *RAI1* mutations alongside known normal controls. Both of these technologies analyze the retention and elution of crude PCR amplimer from a matrix (Kuklin et al. 1997; Kuklin et al. 1999; Li et al. 2002). Those amplimers with no sequence changes (SNPs, insertions/deletions) will theoretically elute as a single homoduplex peak and those with anomalous sequence features will elute as heteroduplex peaks of differing sizes and retention times. In future, we intend to tier our mutation screening by first performing high-throughput DHPLC or TGCE analysis of all PCR products, then sequencing only those products with unusual results. All PCR amplification for DHPLC experiments were performed by Trangenomic, Inc., as our lab had difficulty producing enough crude PCR amplimer for this analysis. As shown in Figures 16 and 17, three known *RAI1* mutations: the 29 bp deletion, 5265delC, and 1449delC were effectively detected by this

Figure 17. DHPLC screening results of known *RAI1* mutations.

Utilizing template DNA and primers from our lab, Transgenomic amplified and analyzed PCR products containing three known mutations in *RAI1*:

Panel **(a)** shows the 29 bp deletion within exon 3, which can be detected at a 1:200 dilution. In panel **(b)** the 5265delC mutation heteroduplex is compared to two normal control samples and the 1449delC mutation heteroduplex is compared to two normal samples in panel **(c)**.

Samples were compared to parental and normal Caucasian control samples using DHPLC analysis. The results of this analysis demonstrate that known mutations can be distinguished from control samples by DHPLC, though the amount of PCR product required for this analysis does not make this method cost effective for our current research purposes.

a.

Absorbance (mV)

UV Detection

Heteroduplex from 29bp del →

Time (minutes)

— SMS129 29 bp del

— 1:20 Dilution of SMS129

— 1:200 Dilution of SMS129

alyzed

1:200
normal
normal

DHPLC
can be
product
current

Figure 17, continued

b.

Absorbance (mV)

Time (minutes)

heteroduplex peak
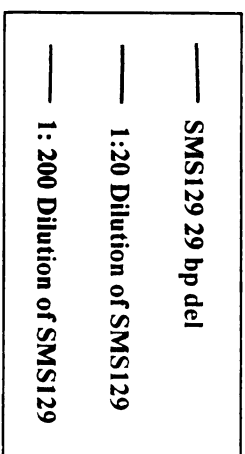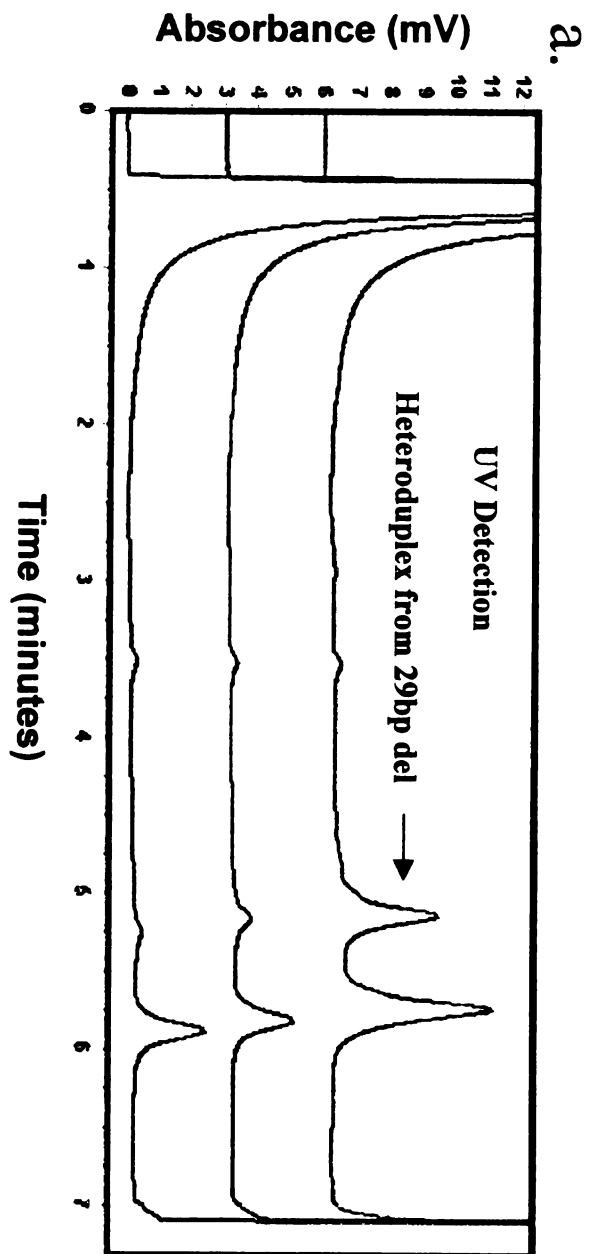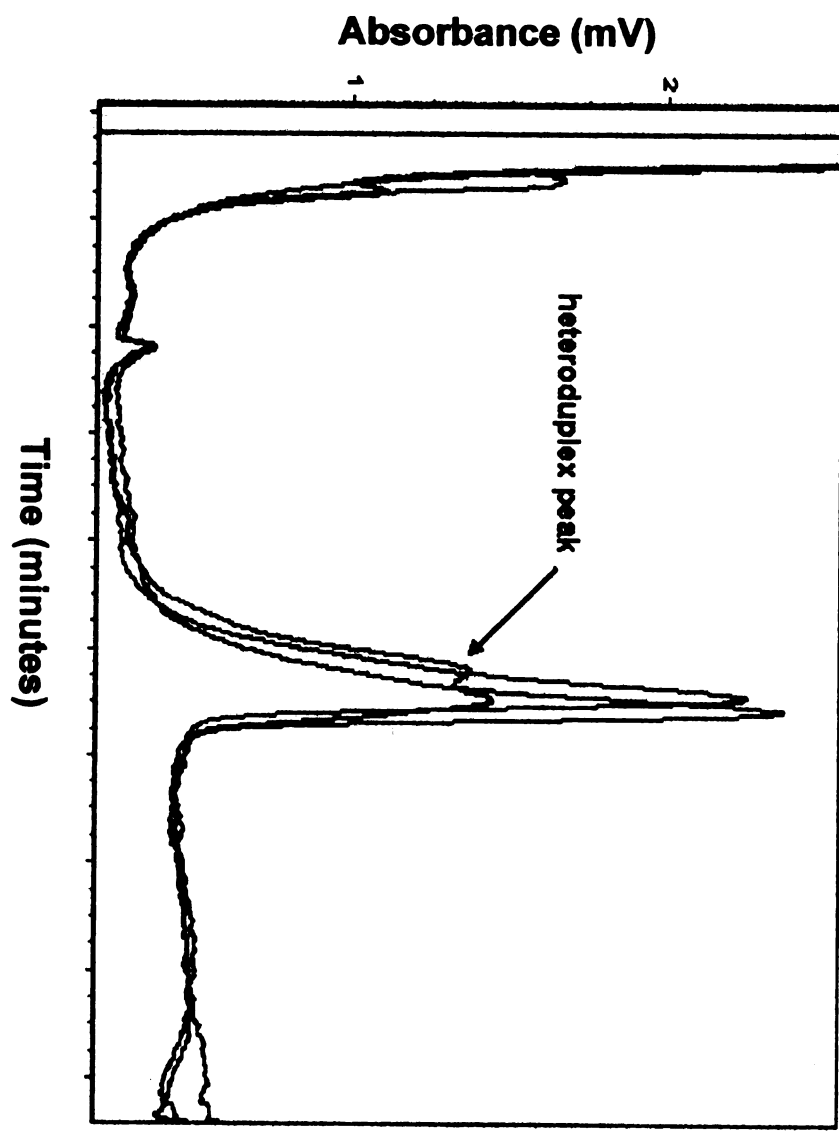
——— Control sample 1
——— SMS156
Heterozygous for 1 bp del
——— Control sample 2

107

Figure 17, continued

C.

heteroduplex peak

Absorbance (mV)

Time (minutes)

Control sample 1
Control sample 2
SMS159
Heterozygous for 1 bp del
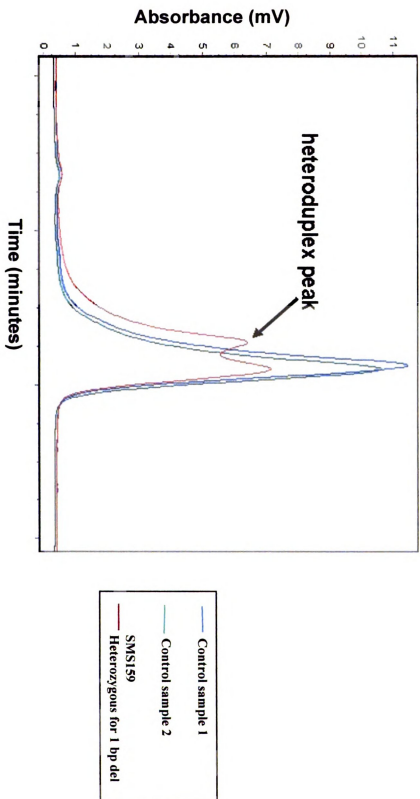
Figure 18. TGCE screening results of known *RAI1* mutations.

In collaboration with Dr. Pragna Patel the Baylor College of Medicine, we analyzed crude PCR products containing three known mutations in *RAI1* using SpectruMedix TGCE analysis: in panel **(a)** the 29 bp del is compared to a normal control, in **(b)** 5265delC is compared to a normal control, and in **(c)** 1449delC is compared to a normal control. As shown in TGCE results, known mutations produce a distinctive heteroduplex peak which can be distinguished from controls.

a.
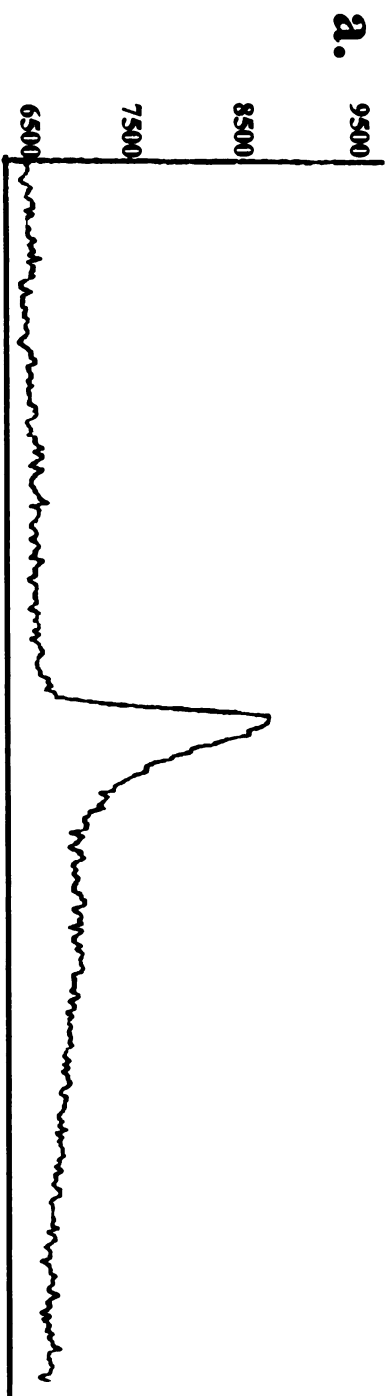
Control sample

SMS129: 29 bp deletion

lyzed
Medix
n (b)
ormal
uplex

110
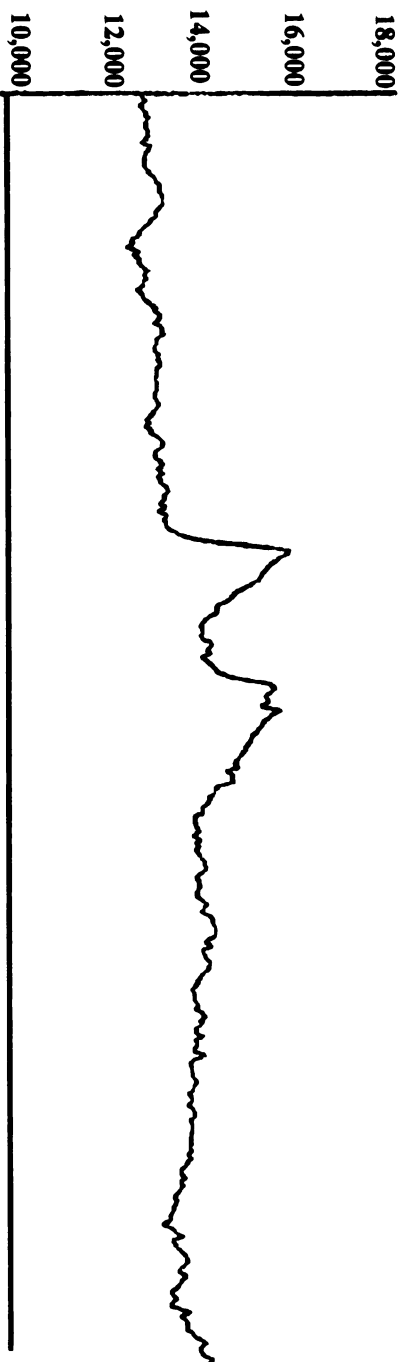
b.

Figure 18, continued
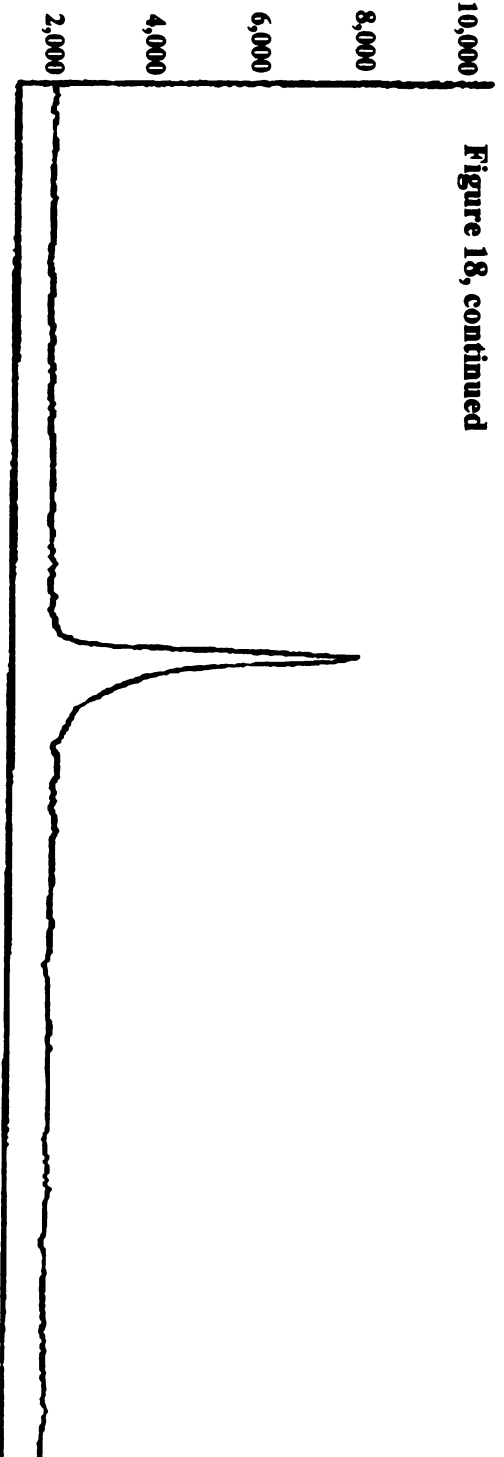
Control sample

SMS156: 5265delC

111
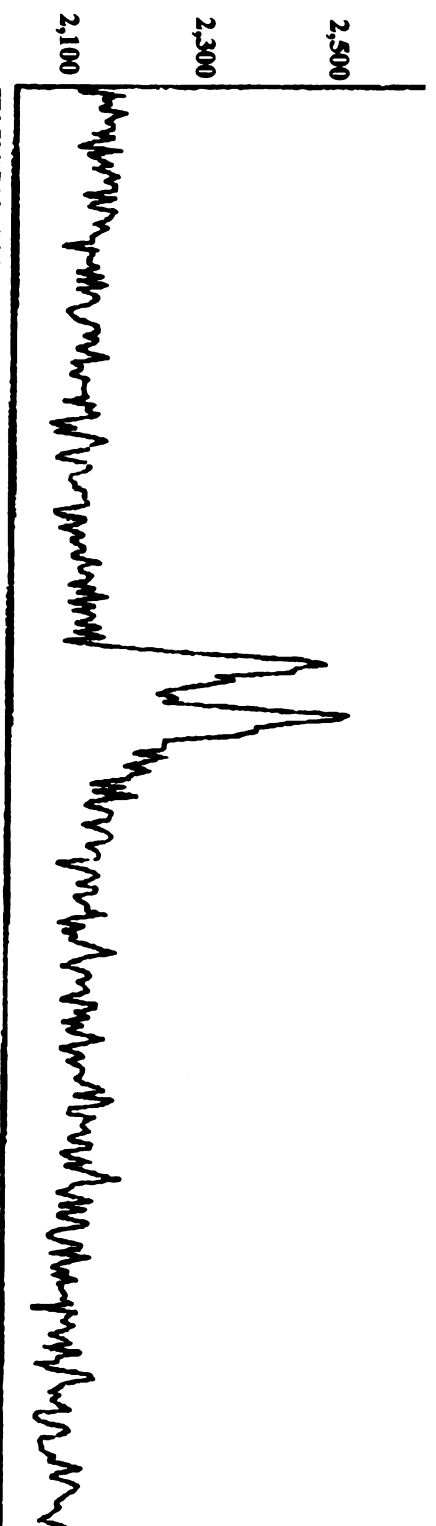
c.

Figure 18, continued
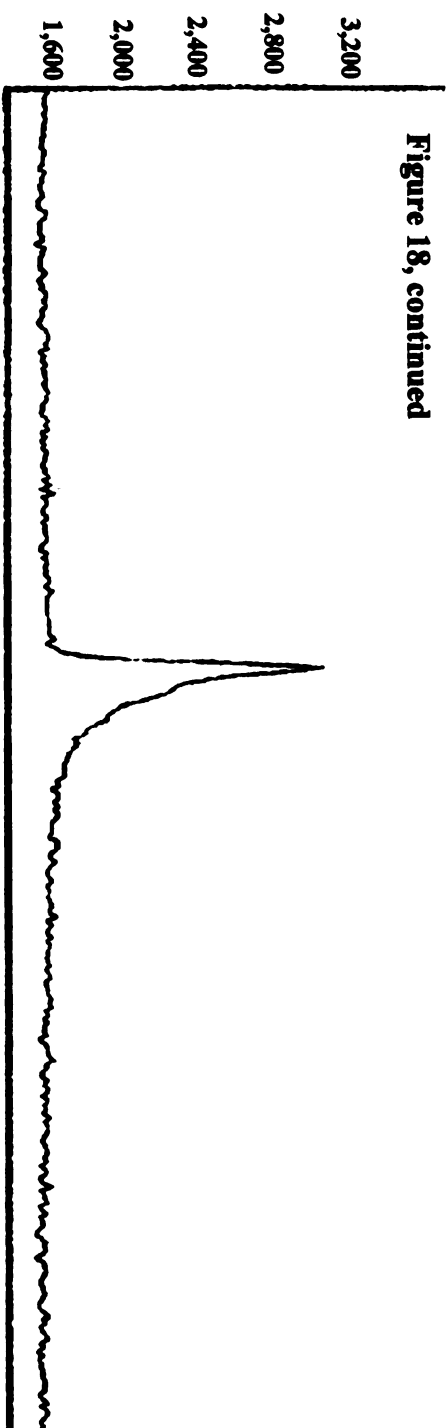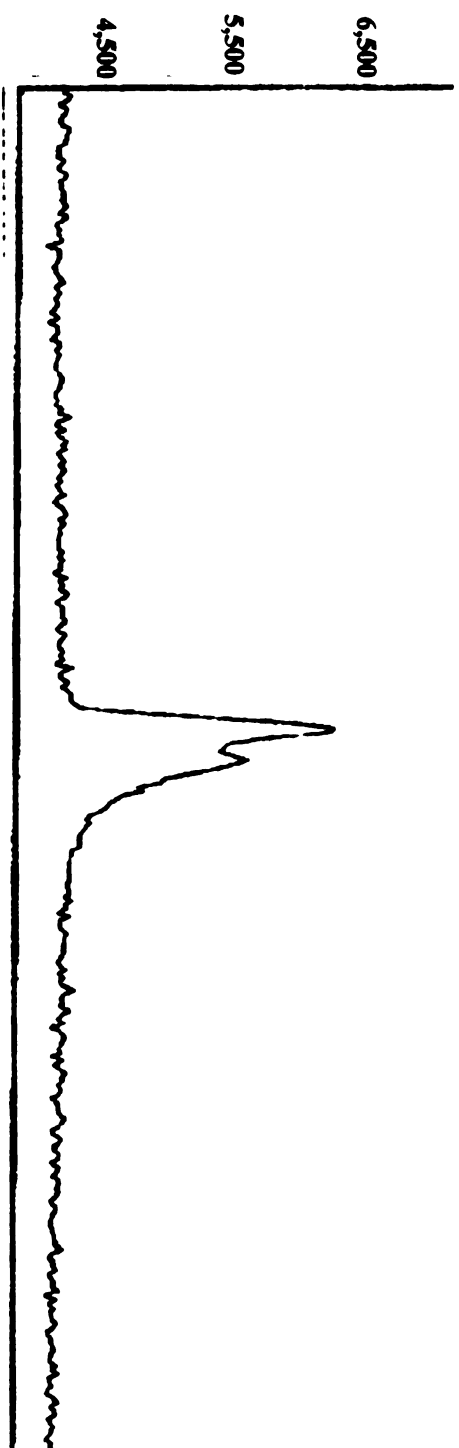
Control sample

SMS159: 1449delC

method, as well as ~20% mosaicism of the 29 bp deletion in the SMS129 parental sample, and a previously undetected SNP from a control sample. While we found DHPLC to be a powerful method for mutation screening, the amount of PCR product required was too great to make this a cost-effective *RAI1* screening method by the Elsea lab.

In contrast, TGCE required less crude PCR product, which Ellen Wilch from the Elsea lab subsequently diluted 1:5 with distilled water and dried overnight before sending these PCR products to collaborators at the Baylor College of Medicine for SpectruMedix TGCE analysis. As shown in Figure 18, our initial results showed that TGCE analysis of the PCR products was able to distinguish known *RAI1* mutations from normal controls, though subsequent analysis with unknown samples did not produce usable results. Possible reasons for this failure are: PCR amplification did not yield enough crude product for analysis, the dilution factor was too high, or the PCR products were too old (optimal analysis is performed with amplimers that are less than a week old, though our initial TGCE samples were at least one month old) (Li, Liu et al. 2002). We hope to optimize our PCR amplification in the future for productive TGCE analysis, though at this time we continue to directly sequence all PCR products.

Conclusions

We believed that *RAI1* was a promising candidate gene for SMS based on preliminary evidence that suggested that it is highly expressed in the brain and may have some functional relationship to a transcriptional coactivator. Though still very little is

known about this large, novel protein. Our approach to evaluating *RAI1* as a candidate gene for SMS was to take advantage of the valuable data supplied by SMS patients with no detectable 17p11.2 deletions. While several clinicians did not truly believe that these individuals had SMS, we began a sequencing effort to identify single-gene mutations in these patients and identified four nondeletion patients who harbored deleterious, frameshift *RAI1* mutations. We now believe that RAI1 haploinsufficiency is responsible for the majority of the SMS phenotype, including the craniofacial, neurological, and ultimately, the behavioral abnormalities, though perhaps not the gross organ defects and cleft palate. Several other SNPs in *RAI1* were found through our mutation screen and at this time, it is difficult to assess the cellular importance of these sequence changes.

Materials and Methods

Patient ascertainment (referred by Brenda Finucane of the Elwyn Institute, Elwyn, PA): This study was approved by the Michigan State University Committee on Research Involving Human Subjects. The phenotypic characteristics discussed here are also summarized in Table 5. SMS129 is a 30 year old man who was admitted to residential placement at age 14 because of aggressive and disruptive behaviors which could no longer be managed at home. He was the product of an uncomplicated full term pregnancy, weighing 7 lbs., 1 oz. Motor milestones were normal but speech development was significantly delayed. From an early age, he exhibited aggressive and self-injurious behaviors, as well as sleep disturbance and frequent "self-hugging". He is obese. His behavior is currently stable, although he continues to require residential placement and

psychotropic medications because of aggression. Results of fragile X and chromosome analyses at age 14 were normal. At age 21, he was reevaluated in genetics clinic and felt to have many behavioral and physical characteristics of SMS. A repeat cytogenetic study (650 band resolution) was normal, as were results of chromosome analysis on skin fibroblasts. SMS156 is a 32 year old obese, white female. Pregnancy and delivery were normal; birth weight was 8 pounds. Motor milestones were achieved on time and there was no history of hypotonia. Speech development was mildly delayed compared to that of her siblings. The patient has a history of mild mental retardation and emotional disturbance, including aggressive and defiant behaviors which prompted residential placement during adolescence. She had self-injurious behaviors, including onychotillomania and polyembolokoilamania, as well as significant sleep disturbance. Her behavior stabilized in the residential setting, and she was able to return home to live with her parents at age 21. She continues to exhibit skin and nail picking, and when excited, the "self-hugging" stereotypy typical of people with SMS. Her facial features are subtly similar to those seen in patients with a 17p11.2 cytogenetic deletion. Chromosome and fragile X analyses at age 17 were normal, as were results of subsequent FISH studies for deletion 17p11.2. SMS159 is a 19 year old patient who has a history of mild mental retardation, self-injury, and aggressive behaviors which resulted in residential placement at age 13. Pregnancy and delivery were uncomplicated, birth weight was 8 lbs. 1 oz. Motor milestones were significantly delayed (walked at 21 months), although he did not have infantile hypotonia. Speech developed within normal limits. Since early childhood, the patient has exhibited "self-hugging" behavior when excited. Macrocephaly was noted in infancy, and MRI studies of his head and spine at

age 11 revealed mild hydrocephalus, Arnold-Chiari malformation, and spina bifida occulta. A temporal lobe cyst was also found and surgically removed. On physical exam at age 13, he was obese and had facial features subtly suggestive of SMS. He had macrocephaly, normal height, gynecomastia and hypogonadism. Results of fragile X and cytogenetic studies were normal, including FISH analysis for deletion 17p11.2. DNA was isolated according to standard protocols from peripheral blood or buccal cells from these patients as well as parental and sibling controls, and Caucasian control samples.

Genomic DNA preparation: In order to isolate template suitable for PCR amplification, DNA was isolated from peripheral blood or buccal cells from all putative SMS patients, parental and sibling controls, and Caucasian control samples.

DNA isolation from whole blood: If $\leq$300 μL of peripheral blood was provided, genomic DNA was extracted using the Puregene kit according to manufacturer's instructions. Briefly, 300 μL of whole blood was mixed with red blood cell lysis solution and incubated for 10 minutes at room temperature. This solution was centrifuged for 20 seconds at maximum speed and the supernatant was removed, leaving the white cell pellet and ~10 μL of residual liquid. The white cell pellet was resuspended and lysed with Puregene Cell Lysis solution and treated with RNase A. The solution was then treated with Puregene Protein Precipitation Solution, vortexed, and centrifuged for 3 minutes at maximum speed. The supernatant containing genomic DNA was removed to a new tube, 300 μL of isopropanol was added, and the mixture was centrifuged at top speed. The DNA pellet was washed with 70% ethanol, dried, and resuspended in 100 μL of distilled water or TE buffer. If larger amounts of blood were supplied (~15-50 mL),

blood was centrifuged at 2000 x g, plasma was removed, and the remaining cells were mixed with solution A (0.32 M sucrose, 10 mM Tris, pH 7.5, 5 mM MgCl$_2$ and 1% Triton X-100) and placed on ice for 30 minutes. The mixture was then centrifuged at 2500 RPM (NEED X G), the supernatant was removed, and 50 mL of solution A were added to the pellet. This solution was placed on ice for 20 minutes, centrifuged as above, and the supernatant was removed. The pellet was resuspended in a solution B consisting of 10 mM Tris, pH 7.5, 400 mM NaCl, and 2 mM EDTA, pH 8 and subsequently digested overnight at 37°C with 100 µL of 20% SDS and 50 µL of 20 mg/mL proteinase K solution. The following day, 3 mL of saturated phenol pH 8.0 were added to the solution while rocking, and then the samples were centrifuged at 2000 x g for 15 mins. The upper phase was removed with a Pasteur pipet and transferred to a new 15 mL polypropylene tube and 3 mL of chloroform:isoamyl alcohol (24:1) were added to this aqueous phase with rocking for 15 mins. Following another centrifugation at 2000 x g, the DNA upper phase was removed and the DNA was precipitated with 2 volumes of 95% ethanol. The DNA was removed with a Pasteur pipet and allowed to sit 70% ethanol for 5 mins, then placed in 200 µL of TE, pH 7.5.

DNA isolation from buccal cells: Genomic DNA was isolated from buccal cells by boiling the cheek brushes in 400 µL of 50 mM NaOH at 95°C for 10 mins. The brush was then discarded and the sample was placed on ice for 10 minutes. This solution was neutralized with 40 µL of 1 M Tris, pH 8.0.

RAI1 *PCR amplification and sequencing reactions*: Analysis of *RAI1* in patient and control DNA was performed by PCR amplification and subsequent sequencing and

analysis of PCR products. PCR primers covering the entire *RAI1* coding sequence, 5' and 3' untranslated regions (UTR) and alternative splice variants (Table 7) were generously provided by Dr. Laura Schmidt of NCI-Frederick or were designed by this laboratory and synthesized at the Michigan State University Macromolecular, Structure, Sequencing, and Synthesis Facility. PCR was performed in a 25 µL volume with 50-200 ng DNA template, essentially as described in Chapter II. PCR amplification was performed in an ABI or MJ Research thermocycler with the following conditions (unless otherwise noted; Table 7): initial denature at 94°C for 4 minutes, 30 cycles of 94°C for 1 minute, 64°C for 1 minute, and 72°C for 1 minutes, and a final extension of 72 °C for 10 minutes. In some cases, Qiagen 5x Q solution or Invitrogen 10x PCR enhancer solution was added to difficult templates. In order to check each PCR amplification, 5 µL of the reaction was electrophoresed in 2% agarose gels containing ethidium bromide. Successful reactions were then purified using the Qiagen Gel Extraction Kit according to manufacturer's instructions or treated enzymatically in the following manner: 2 µL of USB shrimp alkaline phosphatase (1 units/µL) and 1 µL USB exonuclease I (10 units/µL) were added to 5 µL of PCR amplification mixture, the solution was mixed and incubated at 37°C for 15 minutes in a thermocycler, and then inactivated at 80°C for 15 minutes. A sequencing reaction containing at least 10-40 ng of purified PCR product template in distilled water and 30 pmol of sequencing primer (the forward or reverse PCR primer or an internal primer) was then prepared and sequencing was conducted at the Michigan State University Genomics Technology Support Facility using an ABI PRISM® 3100 Genetic Analyzer or ABI PRISM® 3730xl DNA Analyzer.

Sequence analysis: Sequence data from every *RAI1* PCR amplimer was directly compared by BLAST alignments to the published GenBank sequence for *RAI1* mRNA (GenBank AJ271790; AY172136) at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST/) as well as genomic sequence data for the *RAI1* genomic region (GenBank AJ271791; NT_010718).   We searched for putative single nucleotide polymorphisms (SNPs) by searching the NCBI SNP database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp)   or   directly   through   BLAT alignments using the human genome browser (http://genome.ucsc.edu).

Analysis of control samples: In order to determine that specific *RAI1* mutations were not present in relatives of the patients that we had identified as carrying *RAI1* mutations, or in the non-affected Caucasian population, we amplified control genomic DNAs with PCR primers specific to *RAI1* coding exons and intron/exon boundaries (Table 7). The *RAI1* exon 3 29 bp deletion could be detected by digesting the 493 bp exon 3 PCR product with the restriction enzyme *Psp*OMI, which recognizes the G/GGCC sequence present on the non-deleted allele at position 361 of the amplimer, and allele and produces fragments of 361 bp and 132 bp (Figure 11). This restriction enzyme does not cut within the allele carrying the deletion. Following PCR amplification, the samples were digested with 1-2 U of *Psp*OMI at 37°C for 2 hours, then resolved on a 2% agarose gel in 1x TBE containing ethidium bromide.   The resulting fragments were photographed using AlphaImager v5.5 software. The parents and one sibling of SMS129 (Figure 11) were analyzed by this method as well as >100 control samples (data not shown), and none were found to harbor this deletion using this method, though mosaicism was

subsequently detected in one parental sample (SMS128) by Transgenomic DHPLC analysis (Figure 16). In order to assess whether the *RAI1* 5265delC mutation found in patient SMS156 was present in the parents of SMS156 or within the control Caucasian population, >100 DNA samples were screened by PCR amplification and restriction digest with the enzyme *BglI*, which recognizes the sequence GCCNNNN/NGGC beginning at position of the present beginning at position 171 in the wild type allele of the exon 3 amplimer, resulting in fragment sizes of 325 and 171 bp; *BglI* does not cut in the mutated deleted C allele (Figure 12). PCR products were digested with 4-5 U of enzyme, electrophoresed, and photographed as above. Neither the parents of SMS156 (Figure 12) nor the control samples was found to carry the 5265delC mutation. As the 1449delC mutation identified in SMS159 and the 3801delC mutation detected in SMS188 do not alter known restriction sites, in order to assess whether these mutations were polymorphisms, DNA samples from the relatives of SMS159 and SMS188 as well as 100-200 Caucasian control chromosomes were PCR-amplified using exon 3 primers, gel purified or digested with SAP and exonuclease I as described above, and directly sequenced. Neither the parents of SMS159 or the mother and two sisters of SMS188 nor any of the Caucasian control samples were found to harbor these delC deletions.

FISH: FISH was performed by Christopher Vlangos from the Elsea laboratory on patient metaphase chromosomes isolated and prepared using standard cytogenetic protocols. FISH probes were created using BAC or cosmid DNA by using a commercially available nick translation kit to incorporate Spectrum Green or Spectrum Orange dUTP by following manufacturer instructions. Probe DNA (100ng BAC and

180ng cosmid) was precipitated, hybridized to metaphase spreads and washed per manufacturer recommendations (Vysis Inc., Downers Grove, IL). Slides were counterstained using Vectashield antifade with DAPI (Vector Labs, Burlingame, CA). Analysis of the FISH experiments was carried out on a Zeiss Axioplan2 microscope and photographed with a Hamamatsu black and white camera using Zeiss AxioVision software version 2.0.

Northern analysis: In order to determine the embryonic expression pattern of mouse *Rai1*, a SeeGene full-stage mouse conceptus northern blot containing total RNA from mouse embryonic and extra-embryonic tissues was hybridized with ~10^6 cpm/mL of the radioactively-labeled *Rai1* EST clone (IMAGE:1211624) according to the standard northern protocol as described in Chapter II. As an RNA loading control, the housekeeping gene *Gapdh* was subsequently radio-labeled and 10^6 cpm/mL was hybridized to the same blot.

Denaturing High-Performance Liquid Chromatography (DHPLC): DNA templates from SMS129, SMS156, and 159, parental samples and randomly selected Caucasian control samples as well as *RAI1* primer sets were sent to Transgenomic, Inc. All PCR reactions and subsequent DHPLC analysis were performed on-site by Transgenomic personnel.

Temperature Gradient Capillary Electrophoresis (TGCE): Standard PCR reactions were performed as described above by Ellen Wilch in the Elsea laboratory, to amplify known *RAI1* mutations from SMS129, SMS156, and SMS159 template DNA as well as

to amplify the same PCR products from control template DNA. Crude PCR reactions were diluted 1:5 and 1:10 in distilled H_2O, dried overnight at 65°C, and shipped to the laboratory of Dr. Pragna I. Patel at the Baylor College of Medicine. TGCE analysis was performed by Anthony Rohr using Spectrumedix equipment.

Chapter IV.

In vivo evaluation of mouse *Rai1*

Based on our identification of deleterious *RAI1* mutations in several SMS patients with no detectable deletion, we hypothesize that haploinsufficiency of RAI1 is responsible for the majority of the craniofacial and neurobehavioral abnormalities in SMS. In order to assess Rai1 haploinsufficiency in a live animal model, we developed a knockout construct to remove *Rai1* expression. In addition, we assessed *Rai1* dosage sensitivity by constructing BAC transgenic lines and performing preliminary physical assessment of these mice.

SMS mouse models

In order to determine whether *Rai1* dosage-sensitivity could be assessed in a whole animal mouse model, we wished to assess the *in vivo* effects of removal of the Rai1 protein, as well as the effects of overexpression. The research focus of the Elsea laboratory is to create mouse models for specific SMS candidate genes within the deletion interval, as opposed to the large-scale deletion analysis approach of the Lupksi laboratory at the Baylor College of Medicine. This group has created targeted deletions and duplications of a large region of mouse chromosome 11 which has homology of synteny to the SMS common deletion region using Cre-*loxP* targeted recombination chromosomal engineering in 129S5/SvEvBrd ES cells (Walz, Caratini-Rivera et al. 2003). This targeted orthologous chromosomal region contains ~40 genes, most of which are also present in the human SMS common deletion region, although synteny has not

been completely conserved through evolution (Bi, Yan et al. 2002). Specific physical characteristics were evident in the heterozygous deletion [*Df(11)17/+*] and duplication [*Dp(11)17/+*] mice, which were evaluated in a mixed C57BL/6 Tyr^{c-Brd} x 129S5/SvEvBrd background. Specifically, the *Df(11)17/+* mice were significantly obese, had marked craniofacial abnormalities including a shorter snout, were less fertile than control littermates, and some portion of these mice had seizures (Walz, Caratini-Rivera et al. 2003). The duplication mice were noted to be significantly underweight compared to wild type animals (Walz, Caratini-Rivera et al. 2003) Behavioral characterization of these *Df(11)17/+* and *Dp(11)17/+* mice was recently reported (Walz et al. 2004). Open-field locomotor activity studies indicate that male *Df(11)17/+* mice are hypoactive and that male *Dp(11)17/+* male mice are hyperactive (Walz, Spencer et al. 2004). The male duplication mice also demonstrated impaired fear conditioning specific to a context or environment and the deletion mice showed significant circadian rhythm period differences (Walz, Spencer et al. 2004). These Lupski lab mouse studies are an important introduction to assessment of the SMS and dup(17p11.2) phenotype in mice. Several of the physical aspects of SMS have been reproduced in these deletion and duplication mice, demonstrating that haploinsufficiency of the gene or genes which are most important to the syndrome, such as hyperactivity and circadian rhythm abnormalities, can produce measurable effects in an animal model. Further behavioral assessment of the mice may produce more insight into which mental or cognitive pathways have been affected by gene-dosage.

Figure 19. Mouse chromsome 11 genomic region syntenic to the SMS deletion region containing *Rai1* BAC 326M22.

A portion of genomic region of mouse chromosome 11 syntenic to the SMS deletion region is shown in this schematic adapted from the mouse genome browser (http://genome.ucsc.edu). BAC 326M22 (GenBank AC096624), which was used for the Elsea laboratory BAC transgenesis experiments, is highlighted. This BAC contains the entire mouse *Rai1* and *Srebf1* genes and the 3'end of *Tomll2*.

STS Markers

Gap

STS Markers on Genetic and Radiation Hybrid Maps

Gap Locations

bc326W22,

(Tom1/2)

126
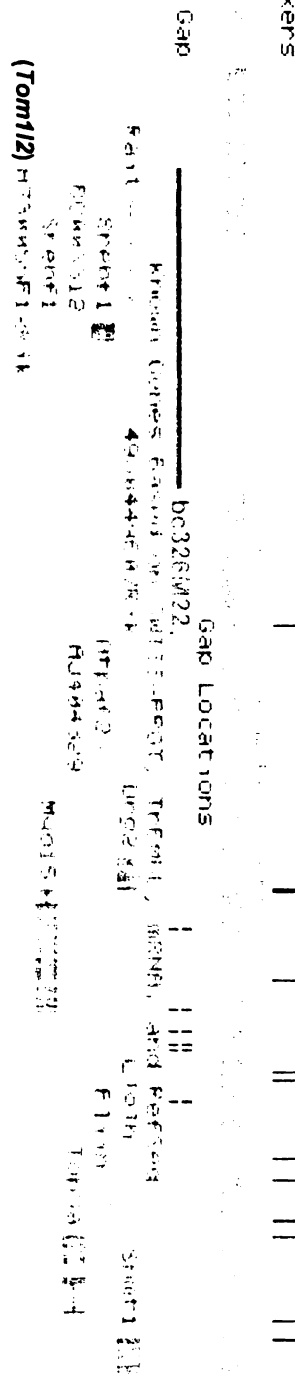
While work using larger deletions as reported above shows that certain aspects of the SMS phenotype can be reproduced in mice, it does not clarify the relationship between specific genes and the phenotype. In order to further define important genes in the SMS region, our lab is interested in creating BAC transgenics and single-gene knockouts in mice. Based on our human *RAI1* mutation screening, we believe that *Rai1* is most likely responsible for the majority of the physical and behavioral abnormalities seen in SMS. Therefore, we have chosen to assess *Rai1* dosage sensitivity in mice by engineering our own gene targeting construct to reduce/eliminate *Rai1* expression *in vivo* as well as to create a stable *Rai1* BAC transgenic lines to determine whether under or overexpression of *Rai1* can reproduce the phenotypic effects of SMS.

BAC DNA isolation

We developed our BAC transgenic mice in collaboration with the University of Michigan Transgenic Animal Model Core (U of M TAMC). Through this arrangement, our lab performed the necessary experiments to isolate and purify the BAC DNA and the TAMC performed the pro-nuclear microinjection, as well as implantation of the potentially transgenic mouse eggs into donor females. The BAC that we chose for injection was RPCI-23 bc326M22 (GenBank accession AC096624), which contains the entire mouse *Rai1* gene as well as the full-length *Srebf1* and the 3' end of *Tom1l2*. This BAC is from a genomic library created from DNA from pooled female C57BL/6J mouse DNA and has been fully sequenced by the mouse genome project. As there is >50 kb of sequence upstream of the *Rai1* translation start site intact within bc326M22, we believe that all of the regulatory elements for correct tissue-specific expression of *Rai1* should be

present in this BAC, although very little is known about the factors that are necessary for the developmental expression of this gene. We chose to use a BAC transgenic rather than a more specific transgenic construct to drive *Rai1* expression because the *Rai1* promoter has not yet been fully characterized, and we may not be able to provide the correct sequences to reproduce the cellular expression pattern. A disadvantage of using the BAC transgenic is that other genes are also present within this relatively large piece (227,682 bp) of genomic DNA. As shown in Figure 19, the entire *Srebf1* is present in this BAC, this gene has been very well-studied and overexpression may not have a measurable phenotypic effect (Shimano et al. 1997), though it remains a possibility that *Srebf1* overexpression may confound our analysis. Since the 5' end of *Tom1l2* is not present in this BAC (Figure 19), we believe that there will not be any expression from this gene. No other genes have been identified in the genomic region present in bc326M22 to date, though small, cryptic genes may exist. It is important to emphasize that the *Rai1* BAC transgenic is a first step towards understanding *Rai1* overexpression and in the future, more specific experiments involving *Rai1* transgenesis may have to be performed to verify our initial results.

Prior to microinjection, our lab isolated bc326M22 BAC DNA using essentially the same modified Qiagen low-copy DNA isolation protocol that we had used to isolate BAC DNA for 17p11.2 mapping studies (Chapter II Materials and Methods). Minor adjustments were made to the final steps of washing and resuspending the bc326M22 DNA to ensure that this DNA was very clean, yet remained intact and resistant to shearing. Following the recommendations of the U of M TAMC, we dissolved the DNA

in a standard microinjection buffer (Schedl et al. 1993), checked the concentration of the

DNA on a UV spectrophotometer and ran the DNA on a agarose gel to check the DNA

quality. We also linearized 1 μg of the DNA with *Not*I and sent the circular and

linearized forms of bc326M22 DNA to U of M. The TAMC ran the digested and

undigested DNA on a pulsed-field gel to confirm the quality of the DNA prep and

readjusted the concentration of the DNA to 0.5-1.0 ng/μL for injection. The BAC DNA

was injected directly into the fertilized mouse eggs from two separate crosses: [C57BL/6

x (C57BL/6 x SJL)F1] and [(C57BL/6 x SJL)F1 x (C57BL/6 x SJL)F1] and then the eggs

were transferred to recipient mothers.

Assessment of *Rai1* BAC transgenic founder mice

Three weeks following the microinjection of the bc326M22 BAC transgene and

transfer of the eggs to donors, 42 live pups (17 females and 25 males) were born. All

mice appeared outwardly "normal" and healthy. As hybrid strains were used for

transgenesis, genes for coat color were allowed to segregate randomly and were not

indicative of transgenic founder mice. Therefore, the 42 mice were agouti, black, white,

or yellow. In order to determine which animals were transgenic, we developed PCR

assays to amplify the specific T7 and Sp6 insert/vector ends of bc326M22. When BAC

DNA integrates into the genome, often a small portion of the bacterial vector sequences

are also integrated into the mouse genome, and these unique insert/vector sequences can

be amplified and distinguished from wild-type mouse genomic sequence. The PCR

reactions we designed could detect the bc326M22 spiked into the mouse genome from a

range of 0.1-100 BAC copy level. An example of the T7 and Sp6 PCR assays are shown

Figure 20. Copy standard and PCR evaluation of *Rail* BAC transgenic mice.

(a) and **(b)** PCR results of the bc326M22 BAC copy standard are demonstrated for the Sp6 and T7 bc326M22 specific insert/vector ends against a wild-type mouse genomic background.   DNA from bc326M22 was spiked into mouse tail DNA at the concentrations indicated below for 0.1x-100x BAC copies:

0.1x copy	= 3.795 pg BAC DNA spiked into 1 μg tail DNA
1x copy	= 37.95 pg BAC DNA spiked into 1 μg tail DNA
10x copy	= 379.5 pg BAC DNA spiked into 1 μg tail DNA
100x copy	= 3795 pg = 3.795 ng BAC DNA spiked into 1 μg tail DNA

BAC DNA was spiked into 1 μg tail DNA and 200 ng of template was used in the Sp6 and T7 specific BAC-end PCR assays. Wild-type mouse genomic DNA with no BAC was used as a negative control and straight bc326M22 DNA was used a postive control. The entire PCR reaction was run on a 2% agarose gel. The PCR reactions shown in **(a)** and **(b)** were quantitative and were performed to demonstrate to the U of M TAMC that our assays were sensitive enough to routinely distinguish transgenic and non-transgenic animals.

(c) The T7 BAC end PCR assay was used to detect the presence of the BAC within mouse genomic DNA from putative founder transgenic mice. DNA from blood samples were used as template and this PCR was non-quantitative. The results of the T7 PCR indicate that 754 does not carry this specific end of the BAC transgene and 753, 755, and 760 are transgenic.

(d) The non-quantitative Sp6 bc326M22 end PCR assay was used to screen putative F1 offspring sired by founder 755. F1 mouse 218 is positive by PCR for the integrated transgenic *Rail* BAC and 213, 214, 215, 216, and 219 are negative.

Quantitative PCR assays

a.

Sp6 bc326M22 end

| Marker |
| 500 bp |
| H₂O |
| Neg. control |
| 100x |
| 10x |
| 1x |
| 0.1x |
| Pos. control |

b.

T7 bc326M22 end

| Marker |
| 500 bp |
| H₂O |
| Neg. control |
| 100x |
| 10x |
| 1x |
| 0.1x |
| Pos. control |

Non-quantitative PCR assays

c.

T7 bc326M22 end

| Marker |
| 500 bp |
| H₂O |
| Neg. control |
| 754-negative |
| 753 |
| 755 |
| 760 |

d.

Sp6 bc326M22 end

| Marker |
| 500 bp |
| H₂O |
| Neg. control |
| 755-male founder |
| 213 |
| 214 |
| 215 |
| 216 |
| 218 |
| 219 |
| Pos. control |
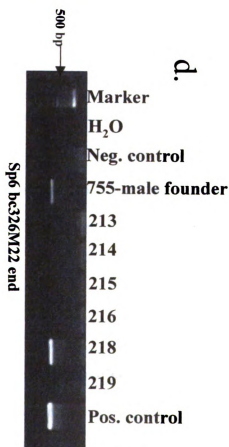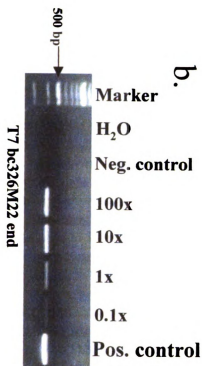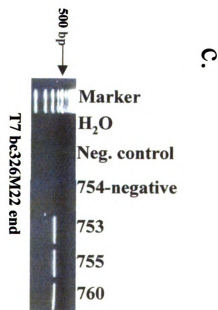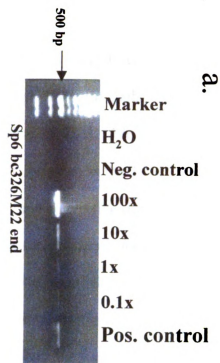
in Figure 20, panels (a) and (b). Initially, at two weeks of age, mouse tails were clipped by the TAMC and shipped to our lab for DNA isolation and identification of the putative transgenic founders. The original DNA isolated from the tails using a phenol-choloroform extraction was not able to be amplified by PCR. We attempted to PCR-amplify the tail DNA with mouse β–globin specific primers, as the TAMC recommends this template as a positive control for DNA quality, and no positive signals were detected. As we were not able to identify the putative transgenic founders before the 42 pups were weaned, all of the mice were transferred to MSU. Peripheral blood was drawn from the foot of all of the animals, and DNA was isolated from whole blood using an alkaline lysis protocol. All of the DNA isolated in this manner from the mice was amplifiable using the mouse β–globin PCR primers. Utilizing the T7 and Sp6 specific primers (Table 9), 9 putative transgenic animals were identified that were positive for the Sp6 bc326M22 end, or both the T7 and Sp6 BAC end. An example of one of the PCR assays used to identify BAC transgenic mice is shown in Figure 20, panel (c). As the template DNA was not quantitated before use in the PCR reactions, the T7 and Sp6 PCR amplifications were not quantitative for bc326M22 copy number. Those animals that were positive for only the Sp6 end may be carrying only a portion of bc326M22. According to the U of M TAMC, it is common to see some portion of BAC DNA which is not fully integrated to the genome.

Copy number determination by Southern analysis

Once several putative transgenic founder mice had been identified by PCR, we attempted to determine the number of copies of bc326M22 that had integrated into the

Primer Name	Forward sequence (5'-3')	Reverse sequence (5'-3')	Annealing temp. (°C)
Rai1 internal fragment	GCCAGCTACAGCAGCTATTCCAA	TCCCGTAAGTGCCTTTGCTATGA	64
Rai1 intronic fragment	CATGCCAGTCCTTCACTATCTG	TGAAGAGAAAAGGATGGTGCAGAG	57
bc326M22 T7 BAC end	CGGTCGGAGCTTGACATTGTAG	GGTAAGAAGTCCCCAGTCC	56
bc326M22 Sp6 BAC end	GATCCTCCCGAATTTCACTAGTG	TCCAGACTTGCTTGTGAACG	58
Mouse β-globin*	CCAATCTGCTCACACAGGATAGAGAGGGCAGG	CCTTGAGGGTGTCCAAGTTGATTCAGGCCATCG	60
Probe *a* to detect *Rai1* gene-targeted homologous recombinants	GAAGCTCTGCCCTTGTTGAG	CTCAACCAGCTTGGCAGAAT	55
Probe *b* to detect *Rai1* gene-targeted homologous recombinants	GCCTACAAATCCAGCTCCAG	TGGCTGGCCTAGATTCATTC	57

Table 9. PCR primers used for amplification of mouse *Rai1*. The specific annealing temperature for each oligo pair is indicated.
*primer sequences from U of M TAMC (www.med.umich/tamc/globin.html)

Figure 21. Southern analysis to determine the integrated *Rail* BAC copy number in founder mice.

In order to assess the copy number of the *Rail* BAC transgenic putative founder mice, 3.5 µg of genomic DNA from isolated from mouse tail was digested with *Hind*III and control mouse genomic standards containing known copy numbers of BAC DNA were also digested as follows: 3.5 µg of control genomic DNA from non-transgenic mice was spiked with BAC DNA representing 1, 3, 5, or 10 copies of bc326M22, just prior to digestion with 4U of *Hind*III. All samples were electrophoresed in 1% agarose gel, transferred to a nylon membrane, and hybridized to radioactively-labeled probes using standard protocols. The probes used to determine the bc326M22 copy number were the 3' end *Rail* cDNA clone (IMAGE:1211624) in **(a)**, the purified bc326M22 Sp6 insert/vector end PCR product in **(b)**, and the purified T7 insert/vector PCR product in **(c)**. The results of copy number calculations for transgenic founder mice from densitometric analysis is tabulated below the autoradiographs. A standard curve was generated from the BAC copy standard hybridizations and putative copy numbers derived from the line equation of the standard curves (and rounded to the next whole number) are indicated in the table for each transgenic founder. The shading indicates that these animals were analyzed on a separate set of Southern blots (data not shown).

a.

Rai1 EST probe

Marker
No bc326M22
1 copy
3 copies
5 copies
10 copies
753
755
760
762
765

b.

Sp6 bc326M22 end

Marker
No bc326M22
1 copy
3 copies
5 copies
10 copies
753
755
760
762
765

BAC-specific
endogenous

c.

T7 bc326M22 end

Marker
No bc326M22
1 copy
3 copies
5 copies
10 copies
753
755
760
762
765

BAC-specific
endogenous
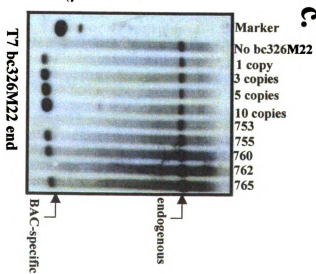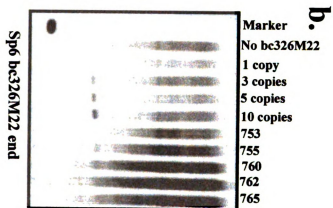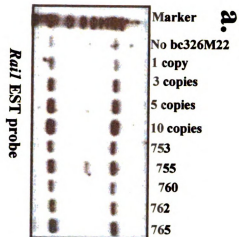
Founder BAC 326M22 transgenic mouse	T7 326M22 BAC end PCR result	Sp6 326M22 BAC end PCR result	Approximate BAC copy number determined by Rai1 EST hybridization	Approximate BAC copy number determined T7 BAC end hybridization	Approximate BAC copy number determined Sp6 BAC end hybridization
753	-	+	3	3	3
755	+	+	4	3	13
760	+	+	2	4	15
762	+	+	3	-	17
765	+	+	5	4	18
767	+	+	1	9	
775	+	+	5	10	4
787	+	+	1		4
792	+	+	13	2	13

135

genome of each mouse. A standard Southern blot hybridization method was developed to compare the positive hybridization signals of the transgenic mice to a standard curve of non-integrated BAC DNA spiked into control mouse genomic at 1, 3, 5 and 10 copies [similar to (Merscher, Funke et al. 2001)]. All DNA was digested with HindIII, which cuts the wild-type Rai1 gene at 2 positions within the coding portion of this gene. For control samples, bc326M22 DNA was spiked into genomic DNA prior to restriction digest. We used DNA probes specific to the 3' end of the endogenous mouse Rai1 (EST clone IMAGE:1211624), as well as the purified bc326M22 insert/vector T7 and Sp6 PCR products to determine the quantitative hybridization signal. As shown in Figure 21, hybridization with the T7 and Sp6 PCR products produced two signals, one specific to the endogenous mouse genomic sequence that is present in the bc326M22 insert and one unique to the BAC end. As predicted by the PCR reactions, not all of the mice were positive for the T7 BAC end and all of the transgenic founder mice were positive for the Sp6 end. Mouse 762, which was positive by PCR for the T7 bc326M22 end, was determined to be negative by hybridization (Figure 21). It is possible that this sample was contaminated during PCR amplification. Densitometry with background subtraction was used to create a standard curve for each Southern blot and the results for were tabulated and graphed using Microsoft Excel, plotting known BAC copy number vs. positive spot density. A putative copy number for each transgenic unknown was then calculated, using the linear equation from the standard curve. Results for the copy number calculations using Southern blots are shown in Figure 21. While Southern blots did confirm that the transgenic founders identified by PCR were actually carrying positive hybridization signal specific to bc326M22, the hybridizations did not produce

clear, unambiguous copy number results. The Sp6 PCR product probe hybridization was particularly difficult to interpret, as this probe contains a mouse-specific repeat and produced a great deal of background hybridization that artificially increased the spot density (Figure 21). Therefore the putative copy number results for the Sp6 hybridization were very high compared to the T7 and *Rail* hybridizations. However, based on the hybridization of the *Rail* EST, which produced the least amount of background signal, some measurable differences in copy number were evident. Founder animals 760, 767, and 775 may have ~1-2 copies of the *Rail* transgenic BAC, animals 753, 755, 762, 765, and 787 may have ~3-5 copies of bc326M22 and animal 792 may have a very high integrated copy number of bc326M22 (Figure 21).

Molecular assessment of *Rail* BAC transgenic F1 mice

Mature male founder *Rail* BAC transgenic mice (755, 760, 762, 767, and 775) were mated to control female C57BL/6 mice purchased from the Jackson Laboratory. The founder males were able to breed normally with the female mice and generally produced litters of ~4-6 F1 generation pups. We began the evaluation of these mice by taking tail biopsies from the 2-3 week old F1 mice and isolating DNA from the tails using the U of M TAMC standard protocol. The tail DNA was first verified using the mouse β–globin PCR reaction, then amplified with the bc326M22 T7 and Sp6 BAC end specific primers (Figure 20, panel d). Three of the transgenic founder males, 755, 760, and 775, produced transgenic F1 offspring; male founders 762 and 767 were not transmitting. Southern analysis of DNA from F1 animals compared to founder DNA (using the same hybridization probes as in Figure 21) demonstrated that the offspring of 760 and 775

137

Rail BAC transgenic founder	Number of litters sired	Total number of F1 offspring	Number of non-transgenic F1 males offspring	Number of transgenic F1 male offspring	Number of non-transgenic F1 female offspring	Number of transgenic F1 female offspring	Percentage of transgenic offspring
755	4	23	5	2	8	8	43%
760	5	27	8	6	7	6	44%
775	2	14	6	4	-	4	57%

Table 10. Summary of F1 *Rail* BAC transgenic breeding from ~August – December 2003. Male founder mice were mated to C57BL/6 females and the resulting F1 mice were analyzed by PCR to determine which offspring were carrying the integrated *Rail* BAC. This table also includes stillborn animals and pups that died shortly after birth and consequently does not accurately reflect the number of animals in Tables 11-14 which were assessed for physical and behavioral anomalies. Two male founders were non-transmitting and their offspring are not represented in this table.

Qualitative assessments:	Non-transgenic F1 males n=14	Transgenic F1 males n=12	Non-transgenic F1 females n=21	Transgenic F1 females n=16
General physical appearance including fur, nails, teeth	Normal	Normal	Normal	Normal
Cage movement	Normal	Normal	Normal	Normal
Righting	Normal	Normal	Normal	Normal
Sound orientation	Normal	Normal	Normal	Normal
Whisker response and ear twitch	Normal	Normal	Normal	Normal
Pupil constriction and dilation including blink	Normal	Normal	Normal	Normal
Sniffing, licking, urination, defecation	Normal	Normal	Normal	Normal
Wild running or freezing	Normal	Normal	Normal	Normal
Postural reflex	Normal	Normal	Normal	Normal
Response to being picked up by tail	Normal	Normal	Normal	Normal
Gait assessment	Normal	Normal	Normal	Normal
Nest building	Normal	Normal	Normal	Normal
Tube test	Normal	Normal	Normal	Normal
Wild running, freezing or unusual movements?	None	None	None	None

Table 11. ***Rai1* BAC transgenic F1 offspring 5 week qualitative assessment data.** For all quantitative assessments, data for male and female transgenic offspring of male transgenic founders were compiled and compared to data from all non-transgenic male and female offspring (as control data, initial assessments were also carried out on the non-transgenic offspring of the non-transmitting males). No obvious differences in transgenic and non-transgenic offspring were noted.

Quantitative assessments:*	Non-transgenic F1 males n=14	Transgenic F1 males n=12	Non-transgenic F1 females n=21	Transgenic F1 females n=16
Cage top hang test	11.8 ± 12.2	6.9 ± 7.8	10.6 ± 9.5	14.2 ± 12.9
Measurements:				
Top of head to tip of tail	21.4 ± 1.7	21.3 ± 1.9	20.7 ± 1.6	19.7 ± 1.1
Length of trunk	64.6 ± 2.9	63.1 ± 2.9	60.9 ± 2.8	59.5 ± 3.5
Distance from outer eyes	11.3 ± 0.5	11.3 ± 0.87	10.9 ± 0.6	10.3 ± 0.7
Distance from outer ears	22.7 ± 3.7	21.8 ± 2.7	23.6 ± 4.8	21.7 ± 3.6
Length of front left limb	from toe to knee: 14.5 ± 2.3 from knee to hip: 11.6 ± 0.8	from toe to knee: 14.7 ± 1.4 from knee to hip: 11.8 ± 0.8	from toe to knee: 13.1 ± 2 from knee to hip: 11.1 ± 0.8	from toe to knee: 13.1 ± 2 from knee to hip: 11.1 ± 0.8
Length of front right limb	from toe to knee: 14.6 ± 2.4 from knee to hip: 11.6 ± 0.9	from toe to knee: 14.3 ± 1.5 from knee to hip: 11.8 ± 1.3	from toe to knee: 13.5 ± 2 from knee to hip: 11.8 ± 1	from toe to knee: 13.4 ± 2 from knee to hip: 11.4 ± 1.4
Length of back left limb	from toe to knee: 14.7 ± 1.3 from knee to hip: 15.1 ± 2.5	from toe to knee: 14.7 ± 0.7 from knee to hip: 14.8 ± 1.7	from toe to knee: 14.8 ± 0.8 from knee to hip: 14.9 ± 1.6	from toe to knee: 14.8 ± 1 from knee to hip: 14.7 ± 1.4
Length of back right limb	from toe to knee: 14.9 ± 1 from knee to hip: 14.6 ± 2.6	from toe to knee: 14.8 ± 0.6 from knee to hip: 14.8 ± 1.7	from toe to knee: 14.9 ± 0.9 from knee to hip: 14.9 ± 1.3	from toe to knee: 15 ± 0.9 from knee to hip: 14.7 ± 1.5

Table 12. *Rai1* BAC transgenic F1 offspring 5 week quantitative assessment data. For all quantitative assessments, average values for male and female transgenic offspring of transgenic founders were compiled and compared to average values from all non-transgenic male and female offspring (as control data, initial assessments were also carried out on the non-transgenic offspring of the non-transmitting males). Detailed analyses of weight and total body length are presented in Figures 23 and 24. *The hot plate assessment is not included in this table as some F1 offspring were initially measured at 55°C instead of 60°C. No obvious differences in transgenic and non-transgenic offspring were noted.

Qualitative assessment:	Non-transgenic F1 males	Transgenic F1 males	Non-transgenic F1 females	Transgenic F1 females
General physical appearance including fur, nails, teeth	n=14	n=12	n=21	n=16
	Normal	Normal	Normal	Normal
Cage movement	Normal	Normal	Normal	Normal
Righting	Normal	Normal	Normal	Normal
Sound orientation	Normal	Normal	Normal	Normal
Whisker response and ear twitch	Normal	Normal	Normal	Normal
Pupil constriction and dilation including blink	Normal	Normal	Normal	Normal
Sniffing, licking, urination, defecation	Normal	Normal	Normal	Normal
Wild running or freezing	Normal	Normal	Normal	Normal
Postural reflex	Normal	Normal	Normal	Normal
Response to being picked up by tail	Normal	Normal	Normal	Normal
Gait assessment	Normal	Normal	Normal	Normal
Nest building	Normal	Normal	Normal	Normal
Tube test	Normal	Normal	Normal	Normal
Wild running, freezing or unusual movements	None	None	None	None
Neurological assessment*: Nest-building	Mice were typically moving, huddled off nestlet, or chewing the nest material	Mice were typically moving, huddled off nestlet, or chewing the nest material	Mice were typically moving, huddled off nestlet, or chewing the nest material	Mice were typically moving, huddled off nestlet, or chewing the nest material

Table 13. *Rai1* **BAC transgenic F1 offspring 10 week qualitative assessment data.** Assessment data for transgenic offspring of the male transgenic founder mice were compiled and compared to data from all non-transgenic offspring (as control data, initial assessments were also carried out on the non-transgenic offspring of the non-transmitting males). No obvious differences between transgenic and non-transgenic offspring were noted. *The tube test was not included in this table, as tubes of various diameters were used for this neurological assessment.

141

Quantitative assessments:	Non-transgenic F1 males	Transgenic F1 males	Non-transgenic F1 females	Transgenic F1 females
	n=14	n=12	n=21	n=16
Top of head to tip of tail	21.7 ± 0.9	21 ± 1	21.1 ± 1.2	20.6 ± 1.1
Length of trunk	71.7 ± 3.9	72.7 ± 5.3	68.5 ± 5.2	67.3 ± 4.4
Distance from outer eyes	11.9 ± 0.6	11.9 ± 0.3	11.6 ± 0.6	11.5 ± 0.6
Distance from outer ears	22.3 ± 1	22.1 ± 1	21.5 ± 1.1	21.4 ± 1.1
Length of front left limb	from toe to knee: 14.7 ± 1.7	from toe to knee: 14.7 ± 1	from toe to knee: 13.7 ± 1.3	from toe to knee: 13.7 ± 1.3
	from knee to hip: 12.1 ± 1	from knee to hip: 12.2 ± 0.6	from knee to hip: 12.5 ± 0.9	from knee to hip: 12.6 ± 1.2
Length of front right limb	from toe to knee: 15.1 ± 0.8	from toe to knee: 14.8 ± 0.7	from toe to knee: 14.3 ± 1	from toe to knee: 14.1 ± 1.2
	from knee to hip: 12.2 ± 0.7	from knee to hip: 12.2 ± 0.7	from knee to hip: 12.5 ± 1.1	from knee to hip: 12.2 ± 0.8
Length of back left limb	from toe to knee: 15.3 ± 0.5	from toe to knee: 15.2 ± 0.4	from toe to knee: 14.9 ± 0.2	from toe to knee: 14.8 ± 0.7
	from knee to hip: 16.7 ± 1.2	from knee to hip: 16.8 ± 1.1	from knee to hip: 16.5 ± 1.1	from knee to hip: 16.1 ± 1.5
Length of back right limb	from toe to knee: 15.3 ± 0.5	from toe to knee: 15.1 ± 0.3	from toe to knee: 15.2 ± 0.8	from toe to knee: 15 ± 0.9
	from knee to hip: 16.3 ± 1.8	from knee to hip: 16.9 ± 1.4	from knee to hip: 16.9 ± 1	from knee to hip: 16.5 ± 1.2

Table 14. *Rail* BAC transgenic F1 offspring 10 week quantitative assessment data. For all quantitative assessments, average values for male and female transgenic offspring of male transgenic founders were compiled and compared to average values from all non-transgenic male and female offspring (as control data, initial assessments were also carried out on the non-transgenic offspring of the non-transmitting males). Detailed analyses of weight and total body length are presented in Figures 23 and 24. No obvious differences in transgenic and non-transgenic offspring were noted.

consistently displayed the same restriction patterns as 760 or 775 (Figure 22), respectively. However several of the offspring of 755 showed hybridization bands not consistent with those in 755 (Figure 22). This suggests that 760 and 775 have one bc326M22 chromosomal integration sites and that these males are stably transmitting the BAC to their offspring (Table 10). In contrast, founder 755 may have multiple integration sites and may not be stably transmitting this BAC to the next generation. Table 10 summarizes our breeding experiments to date.

All of the F1 offspring of all founders appeared normal and healthy. We developed a battery of fairly simple and reproducible physical and behavioral assessments to evaluate the health of F1 bc326M22 offspring. Most of the tests were suggested by the text, *What's wrong with my mouse?: Behavioral phenotyping of transgenic and knockout mice* (Crawley 2000) and a copy of our health assessment form is supplied in Appendix A. Health assessments were performed at 5 and 10 weeks of age and will be ongoing until the F1 mice reach one year of age. Preliminary qualitative and quantitative data are presented in Tables 11-14. All transgenic offspring were evaluated alongside their wild type littermates. The only physical or behavioral differences that were apparent between the bc326M22 transgenic mice and their normal siblings were small differences in weight and total body length (measured from tip of nose to tip of tail) and these data are presented in detail in Figures 23 and 24. The error bars on each graph indicate one standard deviation above or below the mean. We were especially interested in determining whether the BAC transgenic mice might be significantly underweight, as this particular characteristic was noted in the *Dp(11)17/+*

Figure 22. Southern analysis of *Rail* BAC transgenic founders 755 and 775 and F1 offspring.

Southern analysis similar to the experiment depicted in Figure 21 was carried out to determine if each founder male (755, 760, and 775) had the same restriction pattern as offspring sired by each respective male. In order to analyze the restriction patterns, 3.5 µg of genomic DNA from each founder male or F1 offspring was digested with *Hind*III., electrophoresed in 1% agarose gel, transferred to a nylon membrane, and hybridized to radioactively-labeled probes using standard protocols. The control lane included 3.5 µg DNA from non-transgenic mice spiked with 3 copies of bc326M22 just prior to digestion with *Hind*III. In the autoradiograph shown, the *Rail* 3' end cDNA clone (IMAGE:1211624) hybridized to three fragments of DNA from founders 755 and 775 (indicated by red arrows), while the black arrows indicate several inconsistent bands in the 755 F1 offspring 218, 257 and 261. Two of the bands seen in the 755 F1 offspring are also present in the control DNA containing 3 copies of bc326M22, and these may have arisen from partially digested DNA. However, 218, 257, and 261 have multiple inconsistent bands not seen in 755 or the control sample, suggesting that 755 may have multiple BAC integration sites and may not be transmitting the *Rail* BAC transgene in a stable manner. The 775 F1 offspring 252 shows the same restriction pattern as 775. Hybridization patterns from F1 offspring sired by 760 were also similar to 760 (data not shown), suggesting that founder males 760 and 775 were stably transmitting the integrated bc326M22.

5 kb— 11 kb—

3 copies bc326M22

755-founder **male**

218

257

261

292

775-founder **male**

252
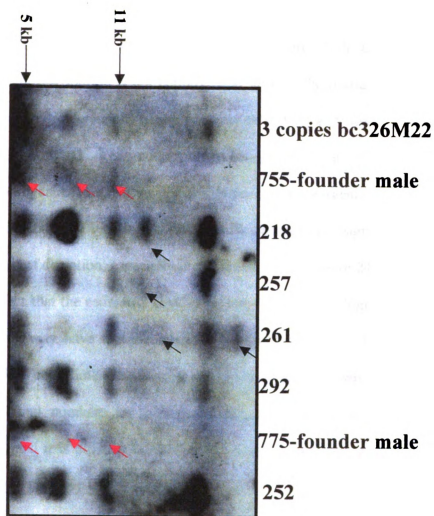
mice (Walz, Caratini-Rivera et al. 2003), though these mice were engineered in a different genetic background. We measured the weights of all transgenic and non-transgenic F1 offspring of 755, 760, and 775 at 5 and 10 weeks. Gender played a significant factor in the weight distribution and females were always statistically underweight compared to males. However, as shown in Figure 23, the effect of the BAC transgene on weight distribution is not entirely clear, though the transgenic offspring of 755 were significantly underweight (using two-way ANOVA) at 5 weeks compared to their normal littermates. There did not appear to be significant interaction between the BAC transgene and gender to affect weight distribution. The transgenic offspring of 755 also had a smaller total body length, though this difference was not significant and there was a high amount of deviation, especially in the female mice (Figure 24). These results may reflect the fact that the estimated BAC copy number in 755 is higher than 760 and 775 or this animal may have mutiple integration sites (Figure 21). Perhaps a higher dosage of *Rai1* does contribute to reduced overall body weight and perhaps body length. However, our animal numbers are low (especially the numbers of transgenic male offspring of 755) and it is difficult to make a statistically significant assessment at this time. When weight vs. body length was plotted in the 10 week old mice (Figure 25), the 755 F1 transgenic and non-transgenic offspring were distributed linearly, while the 760 transgenic and non-transgenic offspring showed similar patterns. It was difficult to assess the 775 offspring as no non-transgenic females were born to this founder. We continue to breed founder males 760 and 775 and will pursue the physical assessment of greater numbers of F1 *Rai1* BAC transgenic offspring.

Figure 23. Weights of *Rail* F1 BAC transgenic animals and normal littermates at 5 and 10 weeks.

(a) The weights of the F1 offspring of each transgenic founder male are shown, comparing non-transgenic males (dark blue) and females (cream; NTM and NTF, respectively) to transgenic males (burgundy) and females (light blue; TM and TF) at 5 weeks. Error bars indicate one standard deviation above or below the mean for each data set. An asterisk (*) indicates statistically significant results as described below ($p < 0.05$).

Statistical analysis of data from F1 offspring of 755: two-way ANOVA, $p=0.009$ for gender, $p=0.0031$ for the effect of transgenic vs. non-transgenic, and $p=0.25$ for interaction between gender and the transgene. Bonferroni posttests showed $p<0.05$ for males* and $p>0.05$ for females; n=2-8.

Statistical analysis of data from F1 offspring of 760: two-way ANOVA, $p=0.003$* for gender, $p=0.85$ for the effect of transgenic vs. non-transgenic, and $p=0.10$ for interaction between gender and the transgene. Bonferroni posttests showed $p>0.05$ for both males and females; n=6-8.

Statistical analysis of data from F1 offspring of 775 (only males could be analyzed as there were no non-transgenic female offspring of 775): unpaired t test, $p=0.22$ for the effect of transgenic vs. non-transgenic; n=4-6.

(b) The weights at 10 weeks of F1 offspring of each transgenic founder male are shown. Total numbers for each subset are indicated and error bars indicate one standard deviation below or above the mean for each data set. An asterisk (*) indicates statistically significant results as described below ($p < 0.05$).

Statistical analysis of data from F1 offspring of 755: two-way ANOVA, $p=0.0014$* for gender, $p=0.099$ for the effect of transgenic vs. non-transgenic, and $p=0.89$ for interaction between gender and the transgene. Bonferroni posttests showed $p>0.05$ for both males and females; n=2-8.

Statistical analysis of data from F1 offspring of 760: two-way ANOVA, $p<0.0001$* for gender, $p=0.75$ for the effect of transgenic vs. non-transgenic, and $p=0.25$ for interaction between gender and the transgene. Bonferroni posttests showed $p>0.05$ for both males and females; n=6-8.

Statistical analysis of data from F1 offspring of 775 (only males could be analyzed as there were no non-transgenic female offspring of 775): unpaired t test, $p=0.35$ for the effect of transgenic vs. non-transgenic; n=4-6.

a.

Raf1 BAC transgenic 5 week weights

Average weight in grams

Transgenic and non-transgenic mice

755   760   775

b.

Raf1 BAC transgenic 10 week weights

Average weight in grams

Transgenic and non-transgenic mice

755   760   775

NTM
TM
NTF
TF

Figure 24. Average total body lengths of *Rail* F1 BAC transgenic animals and normal littermates at 5 and 10 weeks.

(a) The total body lengths (measured from the tip of the nose to the tip of tail) of the F1 offspring of each transgenic founder male are shown, comparing non-transgenic males (NTM, dark blue) and females (NTF, cream) to transgenic males (TM, burgundy) and females (TF, light blue) at 5 weeks. Total numbers for each subset are indicated and error bars indicate one standard deviation below or above the mean for each data set. An asterisk (*) indicates statistically significant results as described below (p< 0.05).

Statistical analysis of data from F1 offspring of 755: two-way ANOVA, p=0.14 for gender, p=0.034* for the effect of transgenic vs. non-transgenic, and p=0.95 for interaction between gender and the transgene. Bonferroni posttests showed p>0.05 for males and females; n=2-8.

Statistical analysis of data from F1 offspring of 760: two-way ANOVA, p<0.0001* for gender, p=0.22 for the effect of transgenic vs. non-transgenic, and p=0.91 for interaction between gender and the transgene. Bonferroni posttests showed p>0.05 for males and females; n=6-8.

Statistical analysis of data from F1 offspring of 775 (only males could be analyzed as there were no non-transgenic female offspring of 775): unpaired t test, p=0.77 for the effect of transgenic vs. non-transgenic; n=4-6.
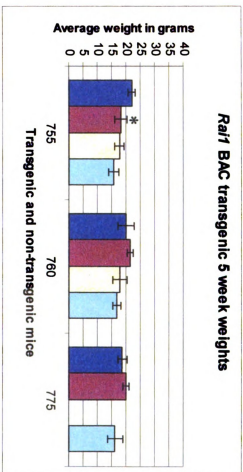
(b) The total body length at 10 weeks of F1 offspring of each transgenic founder male are shown. Total numbers for each subset are indicated and error bars indicate one standard deviation below or above the mean for each data set. An asterisk (*) indicates statistically significant results as described below (p< 0.05).

Statistical analysis of data from F1 offspring of 755: two-way ANOVA, p=0.0032* for gender, p=0.11 for the effect of transgenic vs. non-transgenic, and p=0.92 for interaction between gender and the transgene. Bonferroni posttests showed p>0.05 for males and females; n=2-8.
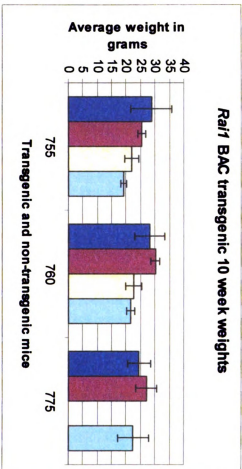
Statistical analysis of data from F1 offspring of 760: two-way ANOVA, p=0.0084* for gender, p=0.77 for the effect of transgenic vs. non-transgenic, and p=0.97 for interaction between gender and the transgene. Bonferroni posttests showed p>0.05 for males and females; n=6-8.

Statistical analysis of data from F1 offspring of 775 (only males could be analyzed as there were no non-transgenic female offspring of 775): unpaired t test, p=0.71 for the effect of transgenic vs. non-transgenic; n=4-6.

a.

Average body length in mm

Raf1 BAC transgenic 5 week total body length

Transgenic and non-transgenic mice

755   760   775

b.

Average body length in mm

Raf1 BAC transgenic 10 week total body length

Transgenic and non-transgenic mice

755   760   775

NTM
TM
NTF
TF
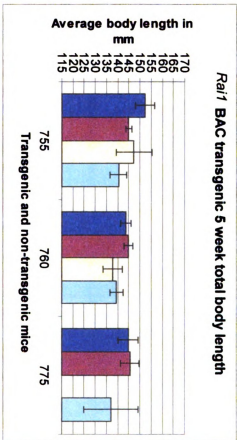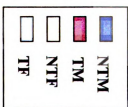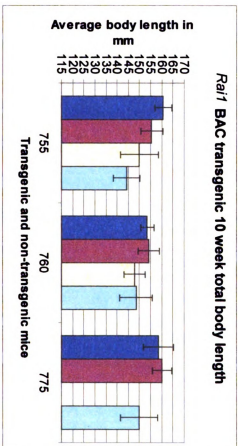
Figure 25. Weight vs. total body length of F1 10 week old *Rail* BAC transgenic mice.

The average 10 weight of F1 offspring is plotted vs. average body lengths for each transgenic founder male are shown. The offspring of 755 shown in **(a)**, the offspring of 760 are shown in **(b)** and the offspring of 775 are shown in panel **(c)**. Each symbol representing a particular dataset is labeled with the genotype (transgenic or non-transgenic) and sex of the animal. There were no non-transgenic female offspring of 775.

a.

755 F1 offspring 10 week weight vs. body length

Average body length in mm

| 140.0 | 145.0 | 150.0 | 155.0 | 160.0 | 165.0 |

Average weights in grams: 0, 10, 20, 30, 40

◆ TF
◆ NTF
◆ TM
◆ NTM

b.

760 F1 offspring 10 week weight vs. body length

Average body length in mm

| 140.0 | 145.0 | 150.0 | 155.0 | 160.0 | 165.0 |

Average weight in grams: 0, 10, 20, 30, 40

◆ TF
◆ NTF
◆ TM
◆ NTM

c.

775 F1 offspring 10 week weight vs. body length

Average length in mm

| 140.0 | 145.0 | 150.0 | 155.0 | 160.0 | 165.0 |

Average weight in grams: 0, 10, 20, 30, 40

◆ TF
◆ NTM
◆ TM

152

At this time, all of the *Rai1* BAC transgenic animals appear to be normal and healthy (Tables 11-14), and only a few animals displayed the characteristic lower body weight that was seen in the *Dp(11)17/+* mice (Walz, Caratini-Rivera et al. 2003). None of the mice appeared to display phenotypes similar to individuals who harbor a duplication of the entire SMS region (Potocki, Chen et al. 2000), though these features can be subtle. However, while we have determined that bc326M22 is present within the genome of these transgenic animals (Figure 21), we have not truly resolved the copy number of each founder or F1 BAC transgenic animals and, most importantly, we have not determined whether there truly is expression from this BAC. Real-time PCR assays were designed to assess copy number, but remain to be optimized and successfully applied using genomic DNA. A summary of these experiments is presented in Appendix B. And, future *Rai1* expression studies may show physiological expression from bc326M22. Also, though we did not observe any outward differences in the F1 mice, we intend to sacrifice ~5 transgenic and non-transgenic animals (males and females) to examine their organs for any gross abnormalities and to perform skeletal evaluations.

Cloning of the *Rai1* knockout vector

In order to create a mouse model of reduced or absent *Rai1* expression, we developed our own specific *Rai1* knockout vector. We cloned two separate targeting arms of homologous sequence from the mouse *Rai1* genomic region into the plasmid targeting vector, pNZTK$_2$ (this plasmid contains a <u>n</u>eomycin resistance cassette for positive selection, the *lac*<u>Z</u> gene, and a <u>t</u>hymidine <u>k</u>inase gene for negative selection). The targeting arms are regions of sequence homologous to the endogenous *Rai1* gene to

allow for correcting targeting and replacement of a large portion of *Rai1*, including the translation start site, with the *lacZ* gene and the *neo* cassette (Figure 26). Similar to the BAC transgenic animal model described above, we performed the molecular biology to create the *Rai1* knockout vector and the U of M TAMC carried out the cell culture and electroporation of the targeting plasmid DNA into Bruce4 C57BL/6 ES cells. The C57BL/6 genetic background will be retained throughout our gene-targeting experiment so we will not have to backcross the animals to a pure genetic background for evaluation. The C57BL/6 genetic background has been used extensively for behavioral evaluations and several standardized studies have been published using this inbred strain. The targeting vector pNZTK$_2$ was chosen for our studies because it contains the β–galactosidase *lacZ* gene, which should replace the endogenous *Rai1* gene (along with the *neo* cassette) if targeting occurs correctly. A map of the pNZTK$_2$ knockout vector and the cloned *Rai1* fragments is shown in Figure 26. Each arm of the targeting vector, the 2.4 kb intronic fragment and the 5.4 kb internal fragment, was amplified by PCR from bc326M22 template using high-fidelity polymerase. Each PCR amplimer was then cloned separately into an Invitrogen TOPO cloning vector, sequenced to confirm, then cut out of the original cloning vectors and ligated into the pNZTK$_2$ cloning vector. A detailed explanation of the cloning steps is provided in the Chapter V. Materials and Methods section. All putative pNZTK$_2$ clones containing both of the *Rai1* gene fragments were sequenced to confirm the correct sequence and orientation. In preparation for microinjection by the U of M TAMC, DNA from the pNZTK$_2$ knockout vector containing both the *Rai1* intronic and internal fragments was isolated using the Qiagen endo-free kit to reduce contamination from endotoxin, and then linearized

Figure 26. *Rai1* **knockout construct.**

A 2.4 kb *Rai1* intronic region fragment was cloned immediately upstream of the start site of *Rai1* translation behind the *lacz* site in the pNZTK2 vector and a 5.4 kb internal sequence representing the 3' end of *Rai1* and the UTR, was cloned into the second polylinker site of the pNZTK$_2$ vector. *Apa*LI (A) and *Eco*RI (E) restriction sites within the *Rai1* genomic region and the pNZTK$_2$ vector are depicted, as well as the probes (*a* and *b*) which will be used in Southern analysis to detect the presence of the endogenous and targeted *Rai1* genes. Following genomic DNA digestion with *Apa*LI and *Eco*RI, probe *a* should anneal to a 5.5 kb fragment in the wild type *Rai1* and 2.8 fragments in the properly targeted gene. Probe *b* should anneal to a 9.2 kb fragment in the endogenous gene and a 7.3 fragments in the recombinant *Rai1*.
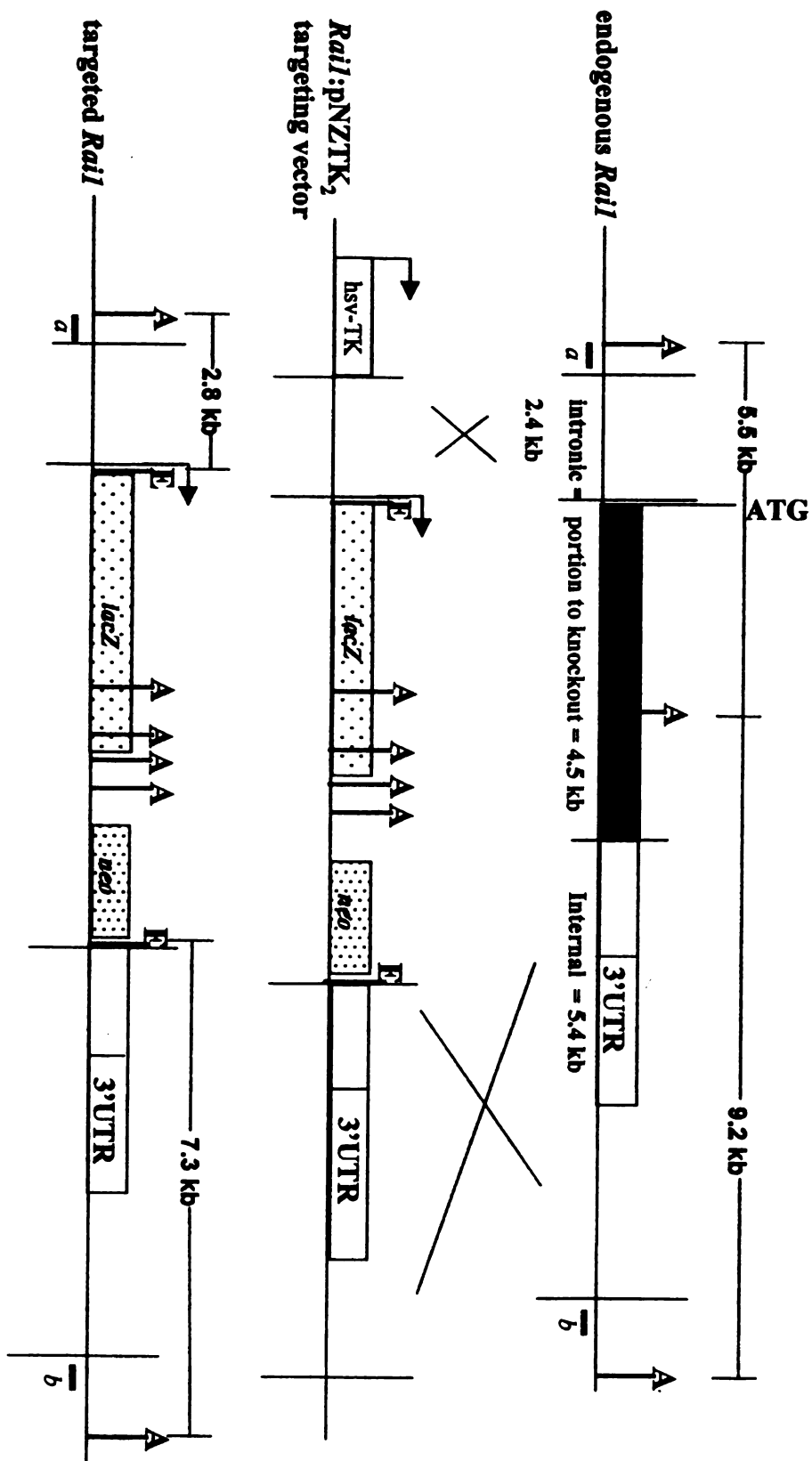
endogenous *Rail*

a

intronic = portion to knockout = 4.5 kb

Internal = 5.4 kb

ATG

3'UTR

5.5 kb

9.2 kb

2.4 kb

b

Rail:pNZTK₂
targeting vector

hsv-TK

lacZ

neo

3'UTR

targeted *Rail*

a

lacZ

neo

3'UTR

2.8 kb

7.3 kb

b

156

utilizing the unique *Asc*I restriction site contained within the vector. The linearized DNA was purified by phenol-chloroform extraction and 200 µg of the *Rail*:pNZTK$_2$ targeting plasmid DNA was sent to the U of M TAMC for electroporation into Bruce4 C57BL/6 ES cells.

Assessment of gene-targeted mice by Southern analysis

As shown in the diagram depicting the *Rail* knockout vector targeting of the endogenous *Rail* gene (Figure 26), there are 3 *Alw*44I/*Apa*LI (these enzymes are isoschizomers) within this genomic region. One specific *Apa*LI site lies within the portion of *Rail* that should be replaced by the *lacZ* and the *neo* cassette from pNZTK$_2$ if the gene-targeting within the ES cells has occurred correctly. Therefore, this *Apa*LI site should be eliminated. A double restriction digest of *Apa*LI and *Eco*RI was used to identify the ES DNA which contains one copy of the endogenous *Rail* and one copy of the *lacZ* and *neo* gene-targeting construct. The new, introduced *Eco*RI sites from the vector polylinker should be present only in gene-targeted DNA. Probes *a* (5') and *b* (3'), which lie outside of the targeting arms have been chosen as probes in Southern blot hybridization (Figures 26, 27).

Following injection of the linearized pNZTK$_2$:*Rail* targeting construct into the Bruce4 C57BL/6 ES cells, the U of M TAMC shipped 5, 96-well plates containing ES cell DNA for screening by our lab to identify homologous recombinants. We digested the ES cell DNA with *Apa*LI and *Eco*RI restriction enzymes and created Southern blots, which were subsequently hybridized with probes *a* and *b*. As shown in the

Figure 27. Southern analysis of ES cell DNA containing the targeted *Rail* knockout construct.

(a) In the 5' probe *a* Southern hybridization, the endogenous 5.2 kb fragment is visualized only in the normal ES cell DNA (represented by plate 3 well E10) and in the targeted ES cell DNA from plate 3 well E13, plate 1 well C12, plate 1 well D9, plate 2 well A3, and plate 3 well H12, both the endogenous 5.5 kb fragment and the targeted 2.8 kb fragment are resolved.

(b) In the 3' probe *b* Southern hybridization, the endogenous 9.2 kb fragment is visualized only in the normal ES cell DNA from plate 3 well E10 and in the targeted ES cell DNA samples from plate 3 well E13, plate 1 well C12, plate 2 well A3, and plate 3 well H12, both the endogenous 9.2 kb fragment and the targeted 7.3 kb fragment are resolved. The probe *b* hybridization from plate 1 well D9 was too faint to discern and is not shown.

a.

Plate 3 E10
Plate 3 E11

endogenous
endogenous
targeted

endogenous
endogenous
targeted

Plate 1 C12

Plate 1 D9

endogenous
targeted

Plate 2 A3
targeted

endogenous
targeted

Plate 3 H12
targeted
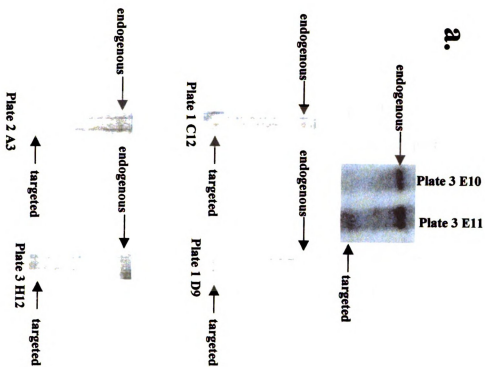
5' probe *a* hybridization

b.

Plate 3 E10
Plate 3 E11

endogenous
targeted

Plate 1 C12

endogenous
targeted

Plate 2 A3
endogenous

Plate 3 H12
endogenous
targeted
targeted
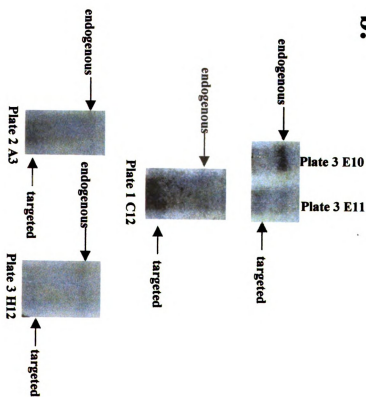
3' probe *b* hybridization

autoradiographs in Figure 27, probes *a* and *b* hybridized to the expected fragments sizes in the endogenous and targeted ES cell DNA. Five putative homologous recombinants were identified by Southern hybridization: plate 1 well D9, plate 3 well E11, plate 1 well C12, plate 2 well A3, and plate 3 well H12. Several other possible homologous recombinants were identified which showed the recombinant fragment with probe *a* but the probe *b* hybridization was ambiguous (often too faint or too much background to clearly distinguish). We also used a probe for the full-length *Rail* gene to detect any rearragements but this probe lies within the targeting vector, so correct targeting to mouse chromosome 11 would not be distinguished from random integration. The ES cells from all possible homologous recombinants will be expanded. More DNA will be harvested from the expansions and the Southern hybridizations with probes *a* and *b* will be repeated for confirmation before the ES cells are injected into albino C57BL/6$^{c2j/c2j}$ blastocysts.

Conclusions

The majority of the molecular biology experiments have been completed to establish stable *Rail* BAC trangenic lines and to inject *Rail* gene-targeted ES cells into blastocysts to produce chimeric mice. The preliminary assessment of the BAC transgenic mice revealed that perhaps higher *Rail* copy number can contribute to a smaller overall body length and less total weight, though the copy number of the these mice needs to confirmed and greater number of animals are required for statistically significance. No other obvious physical or behavioral abnormalities were noted. At least four putative homologous recombinant ES cell lines were identified by Southern hybridization which

carried the targeted *Rai1* construct. These will be confirmed with more ES cell DNA before injection into mouse blastocysts.

Materials and Methods

Mouse copy number calculations: Copy number calculations were determined for RPCI-23 bc326M22 (GenBank Accesssion AC096624) using the copy standards for PCR genotyping protocol devised by the U of M TAMC (http://www.med.umich.edu/tamc/spike.html). For these calculations, the haploid mouse genome is assumed to be 3×10^9 base pairs (bp) and the size of the bc326M22 BAC insert is 227,682 bp. To determine the copy standard for 1 µg of genomic DNA, half of this amount (0.5 µg) is used in the calculations, as transgenic mice are hemizygous:

$$\frac{\text{mass of transgene DNA}}{0.5 \text{ µg genomic DNA}} = \frac{\text{N bp transgene DNA}}{3 \times 10^9 \text{ bp genomic DNA}}$$

$$\frac{\text{mass of transgene DNA}}{0.5 \text{ µg genomic DNA}} = \frac{227,682 \text{ bp insert of transgene DNA}}{3 \times 10^9 \text{ bp genomic DNA}}$$

$$\text{mass of transgene DNA} = \frac{(0.5 \text{ µg genomic DNA}) (227,682 \text{ bp DNA})}{3 \times 10^9 \text{ bp genomic DNA}}$$

$$= 0.00003795 \text{ µg transgene DNA}$$

$$= 37.95 \text{ pg}$$

The copy standard calculations calculated that 1 copy of bc326M22 = 37.95 pg added to 1 µg of genomic DNA. Following this standard:

0.1 copies bc326M22:	3.795 pg spiked into 1 µg genomic DNA
1 copy bc326M22:	37.95 pg spiked into 1 µg genomic DNA
10 copies bc326M22:	379.5 pg spiked into 1 µg genomic DNA
100 copies bc326M22:	3.795 ng spiked into 1 µg genomic DNA

PCR:

<u>Amplification of the mouse *Rai1* genomic region for cloning into the knockout vector</u>:

Template DNA from RPCI-23 bc326M22 was amplified with primers from the genomic

region immediately 5' upstream of the *Rai1* (Table 9) translation start site (hereafter

referred to as the intronic fragment) as well as with primers specific for the central and 3'

end of *Rai1* (hereafter referred to as the internal fragment). These fragments were cloned

into the *Rai1* knockout vector, pNZTK$_2$, (generously provided to the Elsea laboratory by

Dr. Richard Palmiter at the University of Washington) as targeting arms to direct

integration of the pNZTK$_2$ *lacZ* gene and neomycin resistance (*neo'*) cassette into the

mouse genome. The intronic and internal fragments were amplified by standard PCR

protocol (see Chapter II Materials and Methods) using the high-fidelity Accuprime *Pfx*

polymerase (Invitrogen).

<u>T7 and Sp6 PCR</u>: In order to type putative *Rai1* BAC transgenic founder mice, PCR

assays were developed to specifically amplify the RPCI-23 bc326M22 Sp6 and T7

insert/vector BAC ends. As these specific BAC ends are often incorporated into the

mouse genome, the PCR assays were designed to show that we could detect the BAC at

various copy levels within the wild type mouse genomic background (see above for copy

number calculations). Prior to PCR amplification using standard conditions, the

appropriate amount of BAC DNA for 0.1, 1, 10, and 100 bc326M22 copies was spiked

into 1 µg control mouse genomic DNA. Sp6 and T7 bc326M22 end PCR amplification

was carried out using a 200 ng template, utilizing the genomic DNA spiked with BAC

DNA (Figure 20). The entire PCR reaction was electrophoresed on a 2% agarose gel in

1x TBE and visualized. Once we had established a sensitive T7 and Sp6 PCR assay to detect the BAC ends, we assessed the putative *Rai1* BAC transgenic founder mice utilizing template DNA isolated from whole blood using crude alkaline lysis. One μL of DNA from the putative founders was first cycled using primers specific for mouse β-globin to ascertain the DNA quality, then analyzed with bc326M22 insert/vector specific Sp6 and T7 primers.

Cloning of the **Rai1** *knockout vector*: Following amplification of the *Rai1* intronic and internal fragments with Invitrogen Pfx high-fidelity polymerase, 4 μL of the PCR reaction to amplify the internal fragment and 2 μL of the intronic PCR product were separately cloned into the Invitrogen pCR-XL-topo or the pCR-blunt II-TOPO vector, respectively, following manufacturer's instructions. Ligation was allowed to progress for 5 minutes, then 2 μL of the cloning mixture was immediately transformed into TOPO OneShot chemically competent *E. coli* cells and plated onto LB agar containing kanamycin. Colonies formed after overnight growth at 37°C and putative transformants were picked and grown in LB and kanamycin broth culture. These potential transformants were screened by PCR to amplify the polylinker site of the pCR-XL-topo or the pCR-blunt II-TOPO, using –20M13 Universal and M13 reverse primers. Positive TOPO transformants were identified that contained either the *Rai1* 2.4 kb intronic or the 5.4 kb internal fragment.

Once the *Rai1* intron and internal fragments had been amplified and cloned separately into a TOPO vector, plasmid DNA was isolated from the positive clones and

then digested with specific enzymes prior to successive cloning into the pNZTK$_2$ knockout vector. Approximately 0.5 µg of DNA from the *Rai1* internal TOPO clone and 1 µg of DNA from pNZTK$_2$ was digested with *Spe*I and *Not*I. The linearized *Rai1* internal fragment (released from the TOPO vector) and the linearized pNZTK$_2$ DNA were then electrophoresed onto DE-81 ion-exchange paper and purified using the "tombstoning" method. The paper segments containing DNA were microcentrifuged in a 0.65 mL eppendorf tube with a slit in the bottom (placed within a 1.5 mL eppendorf tube to trap the flow) in order to release excess gel running buffer. The 0.65 mL eppendorf tube containing the ion-exchange paper was placed within a fresh 1.5 mL eppendorf tube, and the paper was incubated for 3 minutes in 100 µL high tombstone buffer (1 M LiCl, 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, and 20% ethanol), then centrifuged at room temperature for 3 minutes. Following centrifugation, the eluent was retained; the washing and centrifugation steps were repeated three times. DNA was precipitated from the combined 400 µL of high tombstone buffer by adding 1 mL of 95% ethanol, incubating at −80°C for 15 minutes and centrifuging the mixture at maximum speed for 10 minutes. The purified DNA from pNZTK$_2$ and the *Rai1* internal fragment was resuspended was resuspended in TE buffer. Each fragment was electrophoresed on a 1% agarose gel and the concentration of the DNA was estimated using the Invitrogen 1 kb ladder DNA marker. The internal fragment and the knockout vector were ligated together using T4 DNA ligase at 2:1 insert:vector ratios. The ligation reaction was allowed to proceed for 1 hour at room temperature and the ligation mixture was checked on gel. The ligation then continued overnight at 16°C in a thermocycler. The ligation mixture was diluted 1:5 in 1x TE buffer prior to transformation into Invitrogen DH5α

high-efficiency *E. coli* cells. The transformed cells were then plated on LB agar containing ampicillin. Several individual colonies were picked and grown overnight at 37°C in LB and ampicillin broth cultures. Putative *Rai1* internal:pNZTK$_2$ transformants were screened by PCR using primers specific for pNZTK$_2$ polylinker 2, using 2 µL of overnight bacterial broth culture as template. One colony was isolated that contained the 5.4 kb *Rai1* internal fragment cloned into pNZTK$_2$ in the correct orientation; the insert was also confirmed by direct sequencing.

In order to clone the 2.4 kb *Rai1* intron fragment into the knockout vector, 1 µg of DNA from *Rai1* internal:pNZTK$_2$ transformant was then cleaved by the *Nhe*I restriction enzyme within polylinker 1 and 1 µg DNA from the TOPO cloned intronic fragment was digested with *Nhe*I and *Xba*I. The intronic fragment would be correctly ligated to the *Nhe*I ends of pNZTK$_2$ as the ends of *Nhe*I and *Xba*I are cohesive. The *Nhe*I and *Xba*I digest released the intronic fragment from the TOPO vector. This fragment was electrophoresed on a 1% gel, cut out of the gel, and purified according to manufacturer's instructions using the Qiagen gel extraction kit. The *Nhe*I-digested DNA from the *Rai1* internal:pNZTK$_2$ transformant was electrophoresed through a 1% gel onto DE-81 ion exchange paper, and purified using the "tombstoning" method described above. Similar to the ligation of the larger internal fragment, the linearized intronic and *Rai1* internal:pNZTK$_2$ fragments were electrophoresed on a 1% agarose gel and the concentration of the DNA was estimated using the Invitrogen 1 kb ladder DNA marker. The intronic fragment and the internal:pNZTK$_2$ DNA were ligated together using T4 DNA ligase at 3:1 and 2:1 insert:vector ratios. The ligation reaction progressed for 1 hour at room temperature, after which the ligation mixture was checked on gel. The

ligation then continued overnight at 16°C in a thermocycler. The ligation mixture was subsequently diluted 1:5 in 1x TE buffer before transformation into Invitrogen DH5α high-efficiency *E. coli* cells. The transformed cells were then plated on LB agar containing ampicillin. Several individual colonies were picked and grown overnight at 37°C in LB broth cultures containing ampillicin. Putative pNZTK$_2$ transformants containing the 2.4 kb *RaiI* intron fragment ligated into the *Nhe*I site within polylinker 1 . were identified by PCR using primers specific for pNZTK$_2$ polylinker 1, amplified using 2 μL overnight bacterial broth culture as template. In total, five colonies was isolated that contained the 2.4 kb *RaiI* intron fragment cloned into pNZTK$_2$ in the correct orientation (as well as the previously cloned internal fragment). All inserts were confirmed by direct sequencing.

In preparation for microinjection by the U of M TAMC, DNA from the pNZTK$_2$ knockout vector containing both the *RaiI* intronic and internal fragments was isolated from 250 mL LB and ampicillin broth culture using the Qiagen endo-free kit (in order to reduce contamination from endotoxin), and then linearized utilizing the unique *Asc*I restriction site contained within the vector. The linearized DNA was purified by 1:1 phenol-chloroform extraction, followed by 24:1 chloroform: isoamyl alcohol extraction, and precipitation with 1 M NaCl and 2 volumes of absolute ethanol. The purified DNA was resuspended in 200 μL of sterile 1x TE buffer at a concentration of 1 μg/μL and approximately 200 μg of DNA was sent to the TAMC.

DNA isolation:

BAC DNA: In preparation for microinjection into the fertilized eggs from two separate crosses: [C57BL/6 x (C57BL/6 x SJL)F1] and [(C57BL/6 x SJL)F1 x (C57BL/6 x SJL)F1] for the creation of the *Rai1* BAC transgenic mice, RPCI-23 bc326M22 BAC DNA was isolated as previously described (Chapter II Materials and Methods), except for the following minor modifications to prevent shearing of the BAC DNA. After the elution of the BAC DNA with buffer QF from the Qiagen purification column, the BAC DNA was precipitated with 0.7 volumes of room temperature isopropanol and then centrifuged at ≥ 15,000 x g at 4° C for 30 minutes. The supernatant was then decanted and washed twice with 5 mL of 70% ethanol. The DNA pellet was recovered after centrifugation at 15,000 x g at 4°C for 5 minutes. The ethanol was then removed and the DNA pellet was allowed to dry for 30-60 minutes. We resuspended the pellet in 200 μL of microinjection buffer (10 mM Tris-HCL, 0.1 mM EDTA, pH 8.0, 100 mM NaCl, and a spermine/spermidine polyamine mix), visualized the DNA on a 0.7% agarose gel, stained with ethidium bromide, and quantitated the DNA using an Ultrospec 2000 UV spectrophotometer. Prior to shipping the purified BAC DNA to the U of M TAMC), 1 μg of the bc326M22 DNA was restricted with *Not*I. The U of M TAMC verified the concentration of bc326M22 DNA and checked the quality of the digested and undigested DNA on a pulsed-field gel.

Plasmid DNA isolation: Plasmid DNA from pNZTK$_2$, and Invitrogen pCR-XL-topo or the pCR-blunt II-TOPO was isolated utilizing either the Qiagen mini spin prep or the modified Qiagen midi-prep DNA isolation for greater quanitities of DNA. Plasmid DNA

was isolated from a 5 mL bacterial culture (LB containing appropriate antibiotic), using the Qiagen mini prep kit according to manufacturer's instructions. The protocol for the modified midi-prep was the same as for the cosmid DNA isolation (Chapter II Material and Methods). All plasmid DNA was electrophoresed on a 1% agarose gel and quantitated using an Ultrospec 2000 UV spectrophotometer.

Mouse tail DNA isolation: DNA was isolated from mouse tail biopsies of 0.5-1.0 cm using a standard protocol modified from the U of M TAMC website (www.med.umich.edu/tDNA.html). The tail tips were resuspended in 600 µL of TNES lysis buffer (10 mM Tris, pH 7.5, 400 mM NaCl, 10 mM EDTA, and 0.6% SDS) and 35 µL of 10 mg/mL proteinase K solution and incubated overnight at 55°C. The digested solution was then treated with 166.7 µL of 6 M saturated NaCl, vortexed vigorously for 15 seconds and centrifuged at 14,000 x g for 5 minutes at room temperature. The supernatant was removed to a new tube and the DNA was precipitated with 1 volume of 95% ice cold ethanol. The DNA was pelleted by centrifugation at maximum speed for 10 minutes and dried for 15-30 minutes. The DNA pellet was then resuspended in 100-200 µL of TE (10 mM Tris, pH 8.0, 1 mM EDTA) and allowed to dissolve at room temperature or at 65°C for 10 minutes.

DNA isolation from whole blood for PCR: In order to isolate genomic DNA from whole blood, 500 µL of 1x red blood cell (RBC) lysis buffer was added to 100 µL of whole blood, mixed well and placed on ice for 10 minutes. The solution was then microcentrifuged for 30 seconds at maximum speed and the supernatant was removed

and discarded. The pellet was resuspended in 500 µL of 1x RBC lysis buffer and spun again at maximum speed in a microcentrifuge for 30 seconds. The resulting supernatant was carefully removed and discarded and the RBC lysis was repeated if the pellet remained red. Subsequently 400 µL of 0.05 N NaOH was added to the cell pellet, which was then heated at 95°C for 10 minutes and chilled on ice for 10 minutes. The solution was neutralized with 40 µL of 1M Tris, pH 8.0. This DNA isolation produced DNA appropriate for PCR template though generally did not produce clean enough DNA for Southern analysis.

Sequencing: Direct dye terminator sequencing using ABI PRISM® 3100 Genetic Analyzer, the Applied Biosystems 3730xl DNA Analyzer, or the ABI PRISM® 3700 DNA Analyzer was performed by the Michigan State University Genomics Technology Support Facility (GTSF) to confirm all cloning reactions. Sequencing reactions were prepared according to GTSF requirements (http://genomics.msu.edu/protocols/Custom_DNA_Amounts.html) and consisted of 1 µg of double-stranded template plasmid DNA and 30 pmol of sequencing primer in a volume of 12 µL.

Southern analysis:

Analysis of *Rai1* BAC transgenic mouse copy number: In order to assess the copy number of the *Rai1* BAC transgenic putative founder mice, 3.5 µg of genomic DNA from isolated from mouse tail was digested with 4 U/µg of *Hind*III for ~16-18 hours at 37°C. These genomic DNA digests contained 2.5 mM of spermidine. Standards containing

known copy numbers of BAC DNA were also digested as follows: 3.5 μg of control genomic DNA from non-transgenic mice was spiked with insert μg representing 1, 3, 5, or 10 copies (see above for copy number calculations) of bc326M22 just prior to digestion with 4U of *Hind*III and 2.5 mM of spermidine. All samples were electrophoresed in 1% agarose gel in 1x TAE buffer, transferred to a Hybond-N+ nylon membrane, and hybridized to radioactively-labeled probes using the standard protocol described in Chapter 2. The probes used to determine the bc326M22 copy number were the 3' end *Rai1* EST (IMAGE:1211624), and the purified Sp6 and T7 PCR products representing the specific insert/vector ends of bc326M22.

Analysis of *Rai1* knockout homologous recombinants: In order to identify Bruce4 C57BL/6 ES cell DNA containing the correctly targeted *Rai1*:pNZTK₂ construct integrated within the mouse genome, 3 μg of DNA was digested with *Apa*LI and *Eco*RI restriction enzymes overnight at 37°C. The digested genomic DNA was electrophoresed on a 1% agarose gel in 1x TAE, transferred to Hybond-N+ membrane using standard techniques, and hybridized to radioactively-labeled probes *a* and *b* (Figure 26) as well as a plasmid containing the full-length *Rai1* gene, using the standard protocol described in Chapter II.

Chapter V.

Discussion

The role of *RAI1* in SMS

As this work has described, our molecular analysis of the basis of SMS has produced several promising lines of research. The generation of the physical and transcription map of the critical interval and subsequent EST analysis identified many novel transcripts and several possible candidate genes for the phenotype of SMS. Continued deletion and breakpoint analysis from SMS patients with rare or unusual deletions has enabled our group to further delineate the SMS critical interval. It was especially important for us to take advantage of the valuable information provided by SMS patients with no deletion. While many clinicians doubted that these individuals truly had SMS, we trusted the diagnosis of our collaborators, especially Brenda Finucane of the Elwyn Institute, who had a great deal of practical experience dealing with SMS patients. We were convinced that several of these non-deletion patients displayed characteristic features of SMS, especially the self-abusive and stereotypic behaviors. Utilizing DNA from these putative SMS patients, our lab undertook a brute force mutation screening effort to discover single-gene mutations. We were ultimately rewarded when dominant, deleterious mutations were identified in the *RAI1* in four patients (Table 5, Figures 11-14). This result recasts SMS as a possible single-gene disorder, similar to AGS. Based on the phenotype of those individuals harboring *RAI1* mutations, we believe that RAI1 haploinsufficiency is likely responsible for the physical and neurobehavioral aspects of SMS, which are present in >75% of patients (Table 5).

Other characteristics of SMS which are less common, including large organ abnormalities and cleft palate (Table 5), are likely due to the contribution of other genes within the deletion interval. In regards to the developmental role of other possible SMS candidate genes, our lab is especially interested in *DRG2*, which is extremely conserved through evolution and *TOM1L2*, which may be implicated in membrane trafficking. There are several noteworthy genes that map to the SMS deletion interval within 17p11.2 whose cellular role is currently under investigation.

The true cellular and biochemical role and any possible interacting molecules of the RAI1 protein is also unknown at this time. The RAI1 amino acid sequence contains few clues about possible protein-protein or DNA interactions, except for the putative PHD domain, a nuclear localization site, predicted glycosylation sites, and short regions of sequence similarity to the transcription factor *TCF20* (Rekdal, Sjottem et al. 2000). While the evolutionary relationship between TCF20 and RAI1 proteins is not well-understood, aa analysis demonstrates that both proteins contain PHD (or PHD-like domains in the case of *TCF20*) at their C-terminal ends (Rekdal, Sjottem et al. 2000) (Figure 8). TCF20 contains an ~80 aa long cysteine and histidine-rich zinc finger domain at the C-terminal end of this protein termed a ZNF2 or a PHD/LAP domain, which is closely related to classic PHD domain (Lyngso et al. 2000; Rekdal, Sjottem et al. 2000). Although PHD domains are known to occur in several transcriptional regulators such as trithorax (Aasland, Gibson et al. 1995), their function is poorly understood. In a recent publication by Lyngso et al., this group used different fragments of TCF20 as bait in a yeast two-hybrid experiment and characterized one of the interacting proteins as ring

finger 4 (RNF4). However, only TCF20 constructs not containing the PHD/LAP domain showed positive interaction (Lyngso, Bouteiller et al. 2000). Deletion of this single domain within the C-terminus restored interaction with RNF4 (Lyngso, Bouteiller et al. 2000). This group concludes that the PHD/LAP domain is capable of sponsoring intra-chain interactions within the TCF20 protein and may compete for external zinc finger binding, and therefore acts as a negative regulator of cofactor interaction. Possibly this function occurs in RAI1 and other PHD containing proteins (Lyngso, Bouteiller et al. 2000). It is also interesting to note that TCF20 is not grouped with any of the traditional family of transcription factors (i.e. those with a consensus DNA binding domain), yet it enhances the activity of several well-characterized transcription factors such as c-Jun, Sp1, and Pax6 (Rekdal, Sjottem et al. 2000). Perhaps RAI1 and TCF20 are related members of a family of transactivation proteins.

Future research directions in the Elsea laboratory are focused on *in vitro* and *in vivo* studies of *RAI1* and the mouse homolog *Rai1*, recapitulating some of the TCF20 functional studies (Rekdal, Sjottem et al. 2000). Promising preliminary experiments from the Elsea laboratory as well as IHC experiments performed by collaborators suggest that RAI1/Rai1 is present in the nucleus of mammalian cells (G. Rouleau, personal communication). Subsequent studies are now underway to identify Rai1/RAI1 interacting molecules. Interestingly, the only protein that has shown positive interaction with the mouse Rai1 protein was the zinc finger region of Tcf20, which bound a portion of Rai1 in a yeast two-hybrid experiment (Rekdal, Sjottem et al. 2000). In our research lab, the full-length Rai1 protein will be used as bait to attempt to pull-down other

interacting proteins. It will be useful to also focus on the PHD domain and proteins that may specifically bind this motif, as well as to know if the PHD can sponsor RAI1/Rai1 protein dimerization.

RAI1 mouse models

The *in vitro* analysis described above should provide clues about *Rai1* biochemical pathways. At the same time, detailed phenotypic analysis of the *Rai1* knockout mice may implicate known developmental signalling pathways. This work has focused on the molecular experiments necessary to develop the *Rai1* knockout. Subsequent injection of the recombinant Bruce4 C57BL/6 ES cells (Lemckert et al. 1997) by the U of M TAMC into albino C57BL/6$^{c2j/c2j}$ blastocysts which lack the tyrosinase locus (Schuster-Gossler et al. 2001) should produce *Rai1* germline chimeric animals. The chimeras will be carefully monitored for any obvious abnormal physical or behavioral features and then bred to C57BL/6 females to monitor transmission of the targeted *Rai1* gene. Mice hemizygous and homozygous for *Rai1* disruption will be rigorously assessed with physical and behavioral measurements similar to the battery of tests run on the *Rai1* BAC transgenic mice (Appendix A) and will also be sacrificed and examined internally for gross organ defects. More sophisticated tests such as Morris water maze, open field assessment, curiosity tests, circadian rhythm measurements, and conditioned fear assessments can be applied if the mice appear severely affected, either physically or behaviorally (Crawley 2000). Based on the Lupski lab's targeted disruption of mouse chromosome 11 orthologous to the SMS common deletion interval, *Df(11)17/+* mice heterozygous for this deletion demonstrated marked physical and

behavioral features that mirrored some of those seen in patients with SMS (Walz, Caratini-Rivera et al. 2003; Walz, Spencer et al. 2004). A similar set of experiments to our *Rai1* knockout mice may be performed on mice developed from purchased, targeted mouse ES cells from the German Genetrap Consortium (GGTC). According to the GGTC website (http://genetrap.gsf.de), one 129S2 ES cell line (M073F06; GenBank CL215311) developed by their consortium contains a genetrap insertion between noncoding exon 2 and coding exon 3 (Figure 7), which is designed to sponsor alternative gene splicing into the "trap" instead of the normal *Rai1* mRNA transcript. If genetrap splicing is successful, this *Rai1* genetrap ES cell may create a disruption of the entire gene, though it is possible that leaky splicing "around" the gene trap vector can at low levels (estimated to be ~10% from the GGTC website: genetrap.gsf.de/info/faqs.html#1) may produce an *Rai1* hypomorph. We intend to purchase this ES cell clone for injection into C57BL/6 blastocysts. Mice derived from this method will be in a mixed 129 Sv and C57BL/6 background and may display a slightly different phenotype than our original C57BL/6 knockouts. A disadvantage of using this cell line is the amount of backcrossing required to produce a pure C57BL/6 line for behavioral analysis.

RAI1 and craniofacial development

Craniofacial development is largely controlled by the migration, placement, and expansion of cells derived from the cranial neural crest (Gilbert 2000). Most of the face, including the jaw, teeth, and facial cartilage is a product of neural crest cells. These unique cells, the only neural crest cells to produce cartilage and bone, originate in the hindbrain, which is arranged into compartments called rhombomeres. A gradient of overlapping *Hox* gene expression delineates the rhombomeres and specifies the migration

pattern and fate of the neural crest cells. Cells from different rhombomeres take different migration pathways and are fated to form different tissues; this migration is spatially and temporally controlled (Gilbert 2000). Cells from rhombomeres 1 and 2 travel to the first pharyngeal arch and form the jawbones and the malleus bones of the ear, as well as the bones of the face through a phenomenon of expansion called the frontonasal process (Gilbert 2000). Cells from rhombomere 4 then migrate to the second pharyngeal arch, followed by cells from rhombomere 6, which travel to the third and fourth pharyngeal arch as well as the pouches which will go on to form the thymus, thyroid glands and the parathyroid. The neural crest cells in rhombomeres 3 and 5 join migrating cells which travel on either side of the rhombomere (Gilbert 2000). Other neural crest cells from the fore-and midbrain play a role in the formation of the frontonasal process, palate, and mesenchyme in the first pharyngeal arch. Other cranial neural crest cells originate the mesenchyme that produces the neck bones and muscles (Gilbert 2000).

In the *Rai1* hemizygous and homozygous mutant mice as well our BAC transgenics, it will be important to assess in the mouse embryos whether or not the rhombomere structures and functions are intact. Whole mount immunostaining and *in situ* hybridization will be used to evaluate the gene expression appropriate to various regions of the central nervous system. Several well-characterized markers will be used to visualize the neuroectoderm (*Otx2*), the midbrain/hindbrain boundary (*En-2*), rhombomeres 3 and 5 (*Krox20*), the notochord and floor plate (*Shh*), and rhombomere 4 (*Hoxb-1*) (Alavizadeh et al. 2001). In most cases, an antibody against the appropriate markers can be purchased and used for immunostaining during embryonic days E8.5-

E10.5. Briefly, this procedure involves fixing the embryos overnight in a mixture of methanol/DMSO, then a bleaching step followed by rehydration in phosphate-buffered saline (PBS). The embryos are then treated with a blocking step before the primary antibody is added. Following several washes, the secondary antibody is added and then detected with the appropriate substrate. If the immunostaining yields unsatisfactory results, whole mount *in situ* using DIG-labeled cDNA probes will be used to identify the CNS structures (Alavizadeh, Kiernan et al. 2001). In addition to the assessment of rhombomere structure, we intend to determine if there are any gross abnormalities in the patterning of cranial nerves in the *Rai1* gene-targeted and transgenic mice. Whole mount immunohistochemical analysis with an anti-neurofilament antibody, using the basic protocol described above, will allow for the visualization of cranial nerves (Alavizadeh, Kiernan et al. 2001). In all experiments, normal wild type littermates will be used as controls.

In addition to in-depth embryonic analysis, we will perform craniofacial measurements on newborn and three-month old animals. In their experiments with the heterozygous *Df(11)17/+* mice, the Lupski group found that while newborn mice did not display obvious craniofacial defects, adult mice demonstrated significantly shorter skulls as well as broader and shorter snouts and nasal bones compared to wild type mice (Walz, Caratini-Rivera et al. 2003). We intend to conduct similar measurements for various cranial points, such as the nasal bone, the anterior notch on the frontal process, the intersection of the parietal and intraparietal bones, and the intersection of the interparietal and occipital bones along the midline (Walz, Caratini-Rivera et al. 2003). Measurements

will be conducted on live and sacrificed animals (Walz, Caratini-Rivera et al. 2003).

Skeletons will be prepared similarly to Walz et al. Briefly, cartilage will be stained with

alcian blue and acetic acid, fixed in ethanol, and treated in 2% KOH until clear. The

skeletons will be stained in alizarin red and KOH and cleared with a 1:1 mixture of

ethanol and glycerol. All gene-targeted and transgenic mice will be compared to age-

matched wild type littermates.

Other possible roles of RAI1/Rai1 in development

Another significant physical finding demonstrated in the *Df(11)17/+* mice was

marked obesity (Walz, Caratini-Rivera et al. 2003). As the heterozygous mice harboring

a duplication of the similar chromosomal region [*Dp(11)17/+*] were found to be

significantly underweight, this result suggests that gene dosage effects may influence this

trait (Walz, Caratini-Rivera et al. 2003). We have preliminary evidence that *RAI1* may

influence obesity, as three of the patients harboring deleterious *RAI1* mutations who have

been evaluated by our collaborators (SMS129, SMS156, and SMS159), are all

significantly overweight. Though we currently have a small number of patients, *RAI1*

screening continues in order to identify other individuals with *RAI1* mutations whose

physical traits we can evaluate. As mentioned in Chapter I, the ongoing SMS life history

study at the NIH should provide valuable information about typical body weight change

in SMS patients through adult life. We will also perform frequent weight measurements

on Rai1 knockout mice to determine whether obesity is a trait manifested in these

animals. The Lupski group noted that the both the *Df(11)17/+* and the *Dp(11)17/+* mice

were significantly underweight during the first month of life compared to wild type mice

and that the *Dp(11)17/+* mice remain so throughout their entire lives (Walz, Caratini-Rivera et al. 2003). In contrast, beginning at four months of age, the *Df(11)17/+* mice weigh significantly more that the control animals (Walz, Caratini-Rivera et al. 2003). This finding strongly suggests that *RAI1* or other dosage-sensitive genes within the SMS deletion interval is playing a role in regulating some level of the metabolism controlling overall body weight.

Perhaps the most striking behavioral defects measured in the *Df(11)17/+* mice were the circadian rhythm defects demonstrated by significant period length differences (Walz, Spencer et al. 2004). This rare biological phenomenon directly correlates with the inverted circadian rhythm of melatonin and lack of REM sleep demonstrated by SMS patients (Potocki, Glaze et al. 2000; De Leersnyder, De Blois et al. 2001). Preliminary data also suggests that patient SMS159, who carries a *de novo RAI1* deletion (Figure 13), also has a measurable inversion of the circadian rhythm of melatonin (A. Smith, personal communication). We hypothesize that RAI1 haploinsufficiency has multiple, pleiotropic effects and perhaps may have a global influence on development not only of the cranial neural crest, a metabolic pathway affecting overall body weight, but also the synthesis or secretion of melatonin. We intend to collaborate with Dr. Maya Bucan in the circadian rhythm evaluation of our *Rai1* mouse models. Dr. Bucan's lab has produced elegant studies of the activity and rest cycles of various mutant mice and has devised the proper equipment for these studies (Pickard et al. 1995; Nolan et al. 1997; Kapfhamer et al. 2002). Most of these behavioral studies were completed in the C57BL/6 mouse genetic

background, which is one of the most important reasons that we wished to use C57BL/6

Bruce4 ES cells (Lemckert, Sedgwick et al. 1997) for our *Rai1* targeting project.

Neurological development, retinoic acid and *RAI1*

RAI1 signalling may also have a developmental effect on the neurological system,

which could be extremely sensitive to gene dosage at critical embryonic time points. It is

tempting to link *Rai1/RAI1* to neuronal development simply because of its strong

upregulation following treatment with retinoic acid (RA) (Imai, Suzuki et al. 1995). The

Rouleau group also analyzed *RAI1* cDNA from SKHSH neuroblastoma cells induced

with RA and identified the 5'UTR of *RAI1* (Toulouse, Rochefort et al. 2003). Upstream

of *RAI1* non-coding exons 1 and 2 are significant stretches of CpG dinucleotides, as well

as a RA-responsive element (RARE), which demonstrates that *RAI1* may be directly

inducible by RA (Toulouse, Rochefort et al. 2003).

RA is low-molecular weight, lipophilic substance derived from vitamin A.

Generally, animals take in these retinoid compounds from meat or β–carotene in plants.

Cells acquire RA from the blood system, where RA exists bound to a retinol-binding

protein (Maden et al. 2003). Once inside the cell, retinol is coverted to retinal and then to

RA by retinal dehydrogenases (Maden and Hind 2003). Separate isomers of RA exist,

which subsequently act through different receptors. RA synthesized and modified by the

cell can enter the nucleus and activate or repress gene activity by binding to ligand-

activated nuclear transcription factors (Maden and Hind 2003). In humans as well as

other mammals, there are classes of these transcription factors: retinoic acid receptors

and retinoid X receptors; each class contains several isoforms. These receptors then heterodimerize and recognize RAREs in the promoter regions of RA-responsive genes (Maden and Hind 2003) similar to the sequence found in the 5' upstream region of *RAI1* (Toulouse, Rochefort et al. 2003). It has not been determined whether the RARE in the *RAI1* promoter region has functional significance, though a gel shift assay containing this sequence and various retinoic acid receptors could determine *in vitro* if certain receptors bind to this RARE.

It has been known since the 1950's that RA is highly teratogenic when applied to pregnant animals and can affect multiple organ systems, including the central nervous system (CNS) (Maden 2002). It was subsequently discovered that RA added to cultured embryonal carcinoma cells induced the differentiation of these cells into neurons and glia. RA has the ability to regulate *in vivo* neuronal differentiation through the action of several hundreds of genes, including transcription factors, enzymes, cell-surface receptors, enzymes, and growth factors (Maden 2002). Recent experiments show that RA has an active role in the overall patterning of neurons along the anteroposterior and dorsovental axes (Maden 2002). Signalling by RA seems to be especially crucial for the embryonic development of the hindbrain and anterior spinal cord, where RA acts to position posterior rhombomeres and generate appropriate boundaries (Maden 2002). Another interesting role for RA was demonstrated by Lee et al., who showed that RA and bone morphogenetic proteins have the ability to specify the frontonasal mass and maxillary prominences of chicken embryos. RA levels specifically determine whether certain regions of the face form maxillary or frontonasal forms (Lee et al. 2001). It seems

likely that *RAI1* could play a crucial role in the RA signalling pathway both in developing CNS as well as in craniofacial patterning. It will be very important to examine the hindbrain of *Rai1* heterozygotes and knockout mice for intact expression of various rhombomere markers mentioned above and to assess overall cranial neural patterning. It will also be useful to assess the learning capacity of *Rai1* mutant mice, as mice lacking RA receptors often have severely impaired abilities in the Morris water maze, which measures spatial learning and memory (Maden 2002). Also, very little is known about RA signalling in the adult animal and whether certain adult-onset neurological diseases may be related to dysfunctional RA signalling (Maden 2002). It has been proposed that the etiology of some forms of schizophrenia may be related to RA. Dopamine receptors are regulated by RA and the dopaminergic system is often abnormal in schizophrenia patients (Viggiano et al. 2003). It is interesting that *RAI1* has also been implicated in the etiology of schizophrenia and could be the link between the RA signalling pathway and this complex mental disease.

A possible role for *RAI1* in ADHD and schizophrenia?

As the developmental significance of *RAI1* is still unknown, we can speculate that perhaps other neurological disorders besides SMS that may be caused by perturbations in the amount or quality of the *RAI1* protein. A recent article by Ogdie et al., has reported positive linkage for attention deficit/hyperactivity disorder (ADHD) to a portion of chromosome 17p11 near marker D17S1857 (which maps to a genomic region just distal to the SMS-REPD), using data from 204 nuclear families comprising 270 affected sib pairs (ASPs) (Ogdie et al. 2003). The maximum multipoint LOD score for the 17p11

genomic region was 2.98, which is strongly suggestive for linkage, though the candidate gene region was actually very broad and encompassed 17p11 through 17q11 (Ogdie, Macphie et al. 2003). While not definitively implicating *RAI1* in this genomewide scan, *RAI1* would be a very good ADHD candidate gene based on the hyperactivity and attention-seeking behavior displayed by the four patients analyzed who carry *RAI1* mutations, though the serotonin transporter gene (*5HTT*) maps to 17q11, which several studies have also linked to ADHD (Ogdie, Macphie et al. 2003). It may be possible to screen individuals with ADHD for *RAI1* mutations, though it is difficult to assess the biological significance of missense mutations within the *RAI1* coding sequence, as well as sequence changes within non-coding regions of this gene. As we do not know the true cellular role of the RAI1 protein or the structure of the wild type protein, we cannot determine at this time which changes may have biological significance and which are essentially silent. We could speculate that some form of ADHD may be influenced or caused by hypomorphic *RAI1* mutations or perhaps even a dominant negative effect that an incorrectly functioning RAI1 protein may enact on the normal copy. It is also likely that promoter mutations or those that alter the regulation or the temporal expression of *RAI1* could produce a form of ADHD or even the entire SMS phenotype if the mutation was severe enough. Of course, the genetic background of the individual also plays a crucial role and may further complicate the interpretation of missense mutations. A great deal of research is currently ongoing into genetic factors that contribute to ADHD and most of the emphasis has been on the genes within the dopamine pathway (Sonuga-Barke 2003; Viggiano, Ruocco et al. 2003). *RAI1* may provide insight into an alternative biological pathway of neurological and ultimately behavioral development. This

pathway may contain other particularly dosage sensitive genes which, when altered, can produce ADHD or possibly even SMS.

As discussed in Chapter III, prior to our identification of *RAI1* mutations in SMS patients, many of the publications from the Rouleau group focused on *RAI1* as a possible candidate gene for schizophrenia or perhaps other neurological disorders (Joober, Benkelfat et al. 1999; Hayes, Turecki et al. 2000; Toulouse, Rochefort et al. 2003). These studies were not particularly compelling because they were largely theoretical or statistical and this group offered no biological evidence to support their hypotheses. The Rouleau group was especially interested in the CAG repeats within *RAI1* and whether they may affect the age of onset of SCA2 (Hayes, Turecki et al. 2000) or the association of these repeat regions with responsiveness of certain schizophrenics to conventional neuroleptic treatment (Joober, Benkelfat et al. 1999). No biological confirmation of these conjectures was forthcoming by the Rouleau group or others. But a recent genomewide linkage study involving both 353 ASPs as well as a single pedigree strongly suggests that a locus or loci within 17p11.2-q25 does contribute to schizophrenia (Williams et al. 2003). The maximum LOD score using ASPs for this genomic region was 3.35, though, this candidate region also includes the *5HTT* serotonin transporter gene. However, *RAI1* may also be a priority candidate for detailed sequencing or perhaps haplotype and SNP analysis in this patient population in order to determine whether there may be an association with schizophrenia.

Implications of *RAI1* mutation screening

To date, only 4 of the 12 putative SMS patients we have sequenced were found to harbor deleterious *RAI1* mutations. We cannot definitively determine that all of these non-deletion patients truly have SMS, especially SMS117 and SMS119, who display distinctive, non-SMS characteristics (Table 6). Since our *RAI1* mutation screening has concentrated on the protein coding region and the 3'UTR, it is possible that many of these individuals have deleterious mutations within the RAI1 promoter, regulatory, or enhancer regions. It would be useful to measure the RAI1 protein levels in all putative SMS patients with no deletion compared to parental or wild type sample, to determine if RAI1 is truly haploinsufficient in these individuals. Western analysis of RAI1 prior to mutation screening may be a way to prioritize which nondeletion patients to screen for sequence changes, though we may only be able to easily obtain lymphoblasts for this analysis, which may not show RAI1 expression. But it is also possible that some nondeletion patients may carry a mutation or mutations within a gene in the same developmental pathway that RAI1 plays a role in, or perhaps in a direct signaling target gene of RAI1. Mutations within a single biochemical pathway can produce a similar phenotype in some cases. It is also important to recall that within the cohort of AGS patients without 20p12 deletions, only ~70% were found to carry known *JAG1* mutations (Piccoli and Spinner 2001). Though our pool of SMS patients without 17p11.2 deletions is much smaller than the AGS patients studied, we also have a significant percentage without identified mutations. In both AGS and SMS, it will take time to fully understand which gene or genes and ultimately, which developmental pathways that have been disrupted in these particular individuals. Following our identification of *RAI1* mutations

(Slager et al. 2003), many more nondeletion patients with an SMS phenotype were referred to our group. By sequencing *RAI1* in these individuals, we will obtain a clearer picture of the type of *RAI1* mutations that exist within this population.

Another SMS locus?

In conclusion, we can speculate on the existence of another SMS locus somewhere in the genome. As our cohort of patients with SMS features but no detectable 17p11.2 deletion grows, it is likely that several may be found to carry mutations in *RAI1*, though we may still be left with a number of individuals with no obvious *RAI1* mutation. Perhaps there exists another, etiologically-distinct, genetic syndrome with an extremely similar phenotype to SMS. Or perhaps there is another locus for SMS, which may be associated with a separate gene mutation. As suggested above, once the biological role of *RAI1* is known, other genes in the same biochemical pathway can be screened for mutations. Dominant mutations in these genes may have the same developmental effect as those within *RAI1*. Potentially, *TCF20*, which seems to be evolutionarily related to *RAI1*, could be screened for mutations as well. It is interesting to note that human *TCF20* maps to chromosome 22q13 (Rekdal, Sjottem et al. 2000). Within this particular chromosomal region are at least three known genes which are homologous to genes within the SMS deletion interval. While most of the genes within 22q13 do not have homologous counterparts to those in 17p11.2, *RASD2* (GenBank NM_014310; formerly termed *Rhes*) was localized to 22q13 by sequence analysis, but the original rat *Rhes* was cloned based on its similarity to *RASD1*, which maps to the distal end of the current SMS critical interval (Figure 2). *TOM1L2*, which maps to the middle region of the SMS

critical interval, shares several similar domains with *TOM1* (NM_005488), which maps to 22q13, though *TOM1L2* mRNA expression on a northern blot (Table 1) suggests that this cDNA is at least twice as long as *TOM1*. As sequencing analysis continues, other genes may be identified within 22q13, which may prove to be homologous to SMS candidate genes. We do not have any data in the Elsea laboratory about the possible evolutionary relationship between the 17p11.2 and 22q13 genomic regions and no direct sequence comparisons have been published. As the biological function of these genes is just beginning to emerge, we also cannot determine at this time whether these homologous genes have a truly similar cellular function. It may be unlikely that 22q13 harbors another SMS locus, but as the molecular components of SMS are just beginning to be understood, at this time we cannot rule out any hypothesis. It may also not be a coincidence that positive linkage for bipolar disorder and schizophrenia (Liang et al. 2002) has been mapped to this genomic region as well as 17p11.2.

Conclusions

At this time, the biological implications of identifying mutations in a single gene, *RAI1*, in SMS patients are seemingly limitless. Since so little is known about the cellular role of this gene and whether it is truly involved in RA signaling or mental and craniofacial development, that we can speculate about the various embryonic pathways that could rely on *RAI1* for proper evolution. The global effect of *RAI1* haploinsufficiency could lead to SMS and perhaps missense or hypomorphic mutations could produce ADHD or even schizophrenia. Knowledge of the true biochemical role of *RAI1* awaits the outcome of both *in vivo* and *in vitro* experiments, including the physical

and behavioral assessment of *Rai1* mutant mice. We have helped to advance the molecular analysis of SMS from mapping and gene identification to mutation screening and mice.

APPENDIX A

Elsea lab physical and behavioral assessment for transgenic/gene-targeted mice

Please note that this form usually has entries for mice up to 52 weeks of age, but in the interest of space, only weeks 5-6, 10-12 and 20-22 are shown.

Name of investigator:
Mouse number: **Sex:**
DOB: **Coat color:**
Strain: **Genotype:**

Please note any physical or behavioral abnormalies, even if not specifically indicated on this form

Directions for performing most of the physical and behavioral tests are given at the end of this form

On any given day, please perform the assays requiring anesthesia last!

Physical assessment and simple reflexes	5-6 wks.	10-12 wks.	20-22 wks.
Date assessed			
Weight (grams)			
Whisker appearance			
Skin or fur abnormalities? If so, where?	yes/no	yes/no	yes/no
Bald patches? If so, where?	yes/no	yes/no	yes/no
Condition of genital/rectal areas			
Condition of nails			
Condition of teeth			
Cage movement? Briefly describe	yes/no	yes/no	yes/no
Righting?	Did mouse right itself? yes/no time in secs:	Did mouse right itself? yes/no time in secs:	Did mouse right itself? yes/no time in secs:

Sound orienting?	Right: yes/no Left: yes/no	Right: yes/no Left: yes/no	Right: yes/no Left: yes/no
Pupil constriction/dilation?	Right eye: Constriction? yes/no Dilation? yes/no Left eye: Constriction? yes/no Dilation? yes/no	Right eye: Constriction? yes/no Dilation? yes/no Left eye: Constriction? yes/no Dilation? yes/no	Right eye: Constriction? yes/no Dilation? yes/no Left eye: Constriction? yes/no Dilation? yes/no
Whisker response? Normal/abnormal If abnormal, please comment	Normal/ abnormal	Normal/ abnormal	Normal/ abnormal
Eye blink? If no, please comment	Right eye: yes/no Left eye: yes/no	Right eye: yes/no Left eye: yes/no	Right eye: yes/no Left eye: yes/no
Ear twitch? If no, please comment	Right ear: yes/no Left ear: yes/no	Right ear: yes/no Left ear: yes/no	Right ear: yes/no Left ear: yes/no
Bodily measurements			
Distance from outer ear to outer ear (mm)			
Distance from outer eye to outer eye (mm)			
Distance from top of head to tip of nose (mm)			
Distance from tip of nose to tip of tail (mm)			
Length of trunk from mandible (mm)			
Length of limb -front left (mm)	From toe to knee: From knee to hip:	From toe to knee: From knee to hip:	From toe to knee: From knee to hip:
Length of limb-front right (mm)	From toe to knee: From knee to hip:	From toe to knee: From knee to hip:	From toe to knee: From knee to hip:

Length of limb-back left (mm)	From toe to knee: From knee to hip:	From toe to knee: From knee to hip:	From toe to knee: From knee to hip:
Length of limb-back right (mm)	From toe to knee: From knee to thigh:	From toe to knee: From knee to thigh:	From toe to knee: From knee to thigh:
General behavior-observational assessment			
Wild running? If so, please comment	yes/no	yes/no	yes/no
Freezing? If so, please comment	yes/no	yes/no	yes/no
Sniffing?	yes/no	yes/no	yes/no
Licking?	yes/no	yes/no	yes/no
Rearing?	yes/no	yes/no	yes/no
Defecation?	yes/no	yes/no	yes/no
Urination?	yes/no	yes/no	yes/no
Movement around entire cage?	yes/no	yes/no	yes/no
Sensorimotor reflexes and strength			
Postural reflex Was animal able to stay upright for 10s test? If no, please comment	Cage shake from side to side: Yes/no Cage shake up and down: Yes/no		Cage shake from side to side: Yes/no Cage shake up and down: Yes/no
Response to being picked up by tail Normal/abnormal If abnormal, please comment	Normal/abnormal		Normal/abnormal
Cage top hang test (seconds) Test three times (max 60 seconds)	Test 1: Test 2: Test 3:		Test 1: Test 2: Test 3:

Gaiting test Normal/abnormal If unusual gait, please comment	Normal/ abnormal		Normal/ abnormal
Hot plate test (seconds) Note time in seconds and appropriate assessment (using letters) of mouse's response: a. jump b. raise paw c. paw lick d. paw shake e. other (please comment) Max trial time is 30 seconds Comments:	Time: Response:		Time: Response:

Neurological assessment			
Nest-building assessment At each time point, chose appropriate assessment (using letters) for each nest (more than one can be chosen): a. mouse huddled on nestlet b. mouse huddled off nestlet c. mouse not huddled d. mouse moving e. mouse chewing nestlet f. nestlet chewed and fluffy g. nestlet chewed and flat h. nestlet partially chewed and flat i. nestlet partially chewed and fluffy j. other (please comment) At 120 minutes, measure and record the depth of the thickest part of the nest Comments:		30 min.: 60 min.: 90 min: 120 min.: Depth of nest:	

Tube test Note the mouse numbers for each animal used in test and which one is dominant/submissive as well as whether the animals are siblings or if they have been housed in the same cage	Test 1: Animal numbers: Dominant: Submissive: Animals in same cage? yes/no Test 2: Animal numbers: Dominant: Submissive: Animals in same cage? yes/no Test 3: Animal numbers: Dominant: Submissive: Animals in same cage? yes/no		Test 1: Animal numbers: Dominant: Submissive: Animals in same cage? yes/no Test 2: Animal numbers: Dominant: Submissive: Animals in same cage? yes/no Test 3: Animal numbers: Dominant: Submissive: Animals in same cage? yes/no

Comments:

Has this animal displayed normal breeding behavior?

    \# of litters?

    Size of litters?

Comments:

Cause of death?
    Euthanasia or other_____
    Date of death:
    Age at death:

Following euthanasia:

Skeletal assessment	
Nasal distance (mm)	
Distance between anterior notch on frontal processes (mm)	
Intersection of parietal and interparietal bones (mm)	
Intersection of interparietal and occipital bones (mm)	
Bregma (mm)	
Intersection of maxilla and sphenoid on inferior alveolar ridge (mm)	
Gross examination of internal organs	
Abnormalities?	

Description of physical tests

Righting:	turn the mouse onto its back and measure time in seconds for mouse to right itself; comment if mouse is unable to right itself or unusual response is noted
Sound orienting:	make brief, sharp sound to the right and left of mouse; note if mouse turns head in appropriate direction
Pupil constriction/dilation:	with a pen-light, shine a beam of light in the direction of the mouse's eye; constriction should occur when light is shone and dilation should occur when light is removed
Whisker response:	lightly brush the whiskers of animal with a small paint brush and note response (normal mice will stop moving their whiskers and may turn head)
Eye blink:	approach the eye with the tip of a clean cotton swab and note if eye blinks (test both eyes)
Ear twitch:	touch with ear with the tip of a clean cotton swab and note if ear twitches (test both ears)

Bodily measurements

In a fume hood, use ~ 300 μL of isofluorane to anesthetize the mouse and wait until the animal is fully asleep. Measure the animal according to the diagram and record measurements in mm. Monitor the animal until he/she wakes up and is functioning normally.

Sensorimotor and reflexes

Postural reflex:	place mouse in empty cage shake from side to side and up and down for 10 seconds each; note mouse's ability to maintain upright position
Response to being picked up:	note response when mouse is picked up by tail for 10 s; a normal response is to raise head, extend extremities and reach for ground when lowered
Cage top hang test:	hang mouse in empty cage with a cage top; measure time in seconds for mouse to remain hanging and note crawling movement along cage top; repeat test three times (max 60 seconds)
Gaiting test:	gaiting is measured by coloring each mouse's foot with one color (using non-toxic materials) and walking mouse through a small tunnel on white paper
Hot-plate test:	place mouse on analgesic hot plate set at 60°C and measure time in seconds for mouse to display a common response (jump, raise paw, paw lick or paw shake) or an unusual response (note and comment); max trial time is 30 secs.

Behavioral tests

Nest building assessment: nesting material is placed in the cage with the mouse and the quality and depth of the nest over various time periods (30, 60, 90, and 120 minutes) is assessed (mouse huddled on nestlet, mouse huddled off nestlet, mouse not huddled, mouse moving, mouse chewing nestlet, nestlet chewed and fluffy, nestlet chewed and flat, nestlet partially chewed and flat, nestlet partially chewed and fluffy); at 120 min., the thickest part of the nest is measured

Tube test: two mice (note mouse numbers and whether animals are siblings as well if they have been housed in same cage) are placed in 30 cm x 3.5 cm tube and monitored for dominance (animal pushes the other mouse out of the tube) and submission. The test is repeated with transgenic/non-transgenic mice and with mice from different cages.

APPENDIX B

Preliminary real-time PCR experiments to establish *Rai1* BAC transgenic copy number

As the genomic Southern hybridization shown in Figure 21 did not produce unambiguous *Rai1* BAC copy number results, an alternative method to determine BAC copy number number in our male founder mice as well as the F1 offspring was devised by the Elsea laboratory utilizing Sybr Green real-time quantitative PCR. Our research plan was to compare the PCR amplification of transgenic animals to non-transgenic littermates and to determine DNA copy number by measuring the quantitative PCR results of transgenic animals against a standard curve of known BAC copy number spiked into mouse genomic DNA, similar to the standard curve in Figure 21. The fluorescent molecule Sybr Green binds to double-stranded DNA and is provided as part of a 2x ABI Master Mix solution. The appropriate primers, template DNA and water are added to this mix before real-time amplification. The ABI 7700 (available through the MSU GTSF) detects fluorescent incorporation of Sybr Green during the logarithmic phase of PCR. Three sets of primers were designed for our experiments by the ABI Primer Express program, which should be amplifiable under real-time PCR cycling conditions: 95°C for 5 minutes, and 40 cycles of 95°C for 15 sec and 60°C for 1 minute. Three primer sets were utilized to confirm that the copy number results were the same for each set and to distinguish animals who may not have an intact BAC transgene. The primers (listed below) were designed to amplify a portion of the mouse *Rai1* gene, as well as genomic regions near the bc326M22 Sp6 and T7 BAC ends (note that these

primers contain unique BAC vector sequences in contrast to the T7 and Sp6 primers

listed in Table 9):

Rail F primer
5'-GCCTGAAATCCGACTCAAATACAT-3'

Rail R primer
5'-TGGTGTAAGCATCTCGCTTCT C-3'

Genomic region near Sp6 BAC end F primer
5'-GGTAACCTGGAACCGAGACAT C-3'

Genomic region near Sp6 BAC end R primer
5'-AAGGAAACAAGCCCCAGAAAC-3'

Genomic region near T7 BAC end F primer
5'-TCCCACTAAGGGAGCTCTTTATACC-3'

Genomic region near T7 BAC end R primer
5'-TCCAAATGTCCTTTAGCCCTTTC-3'

The primers were optimized on the Elsea laboratory ABI standard thermocycler,

using the same cycling conditions as the ABI 7700 real-time thermocycler provided by

the MSU GTSF. BAC 326M22 DNA was used to optimize the primers, as genomic

DNA isolated from mouse tails did not provide consistent amplification. All three of the

primer sets amplified the purified BAC DNA. In order to test the Sp6 primers in the real-

time thermocycler, a standard curve of BAC DNA ranging from 37.9 pg-158 ng of BAC

DNA was serially diluted. The reaction was run with a Sp6 final concentration of 300

nM, which was recommended by Dr. Annette Thelen of the GTSF and all samples were

run in triplicate along with a no template control. The results of the standard curve are

shown in Figure 28 which plots the log of the BAC DNA concentration against the

threshold value (C_T). The threshold value is an experimentally determined value

measured during the logarithmic cycle of PCR approximately 10 standard deviations

above a baseline fluorescent value. In the experiment depicted in Figure 28, the threshold

Figure 28. BAC 326M22 DNA quantitative PCR standard curve.

A standard curve of 37.9 pg-158 ng of BAC 326M22 DNA was amplified with primers covering the genomic region near the Sp6 bc326M22 end using the ABI 7700 real-time PCR thermocycler under standard conditions. For this reaction, the baseline value was measured from cycles 3 to 8 and threshold values (C_T) were measured at 0.534. In the graph the log quantity is plotted against C_T. Two of the replicates did not amplify due to human error. The line equation for the standard curve y = -2.8143x + 15.972 and the R^2 value was 0.9814. This experiment demonstrated that quantitative PCR did work in the laboratory on purified BAC DNA, though subsequent experiments with mouse genomic DNA failed.

BAC 326M22 DNA standard curve

$y = -2.8143x + 15.972$

$R^2 = 0.9814$

C_T

Log Quantity
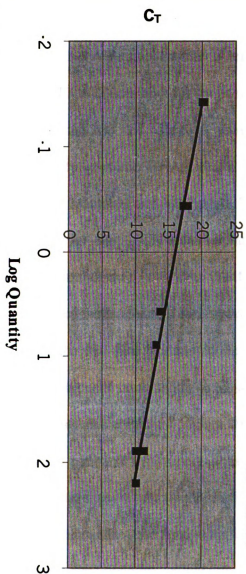
was set at 0.534 and the baseline was measured from cycles 3 to 8. Due to human error, two of the replicates produced no amplification and were not plotted on the graph. The line equation and the R^2 value for the standard curve are displayed on the graph in Figure 28.

The standard curve with BAC demonstrated that the real-time PCR did work with at least one set of real-time primers designed in the Elsea laboratory and subsequent experiments with the *Rai1* and T7 PCR sets also shown positive amplification with BAC DNA at a primer concentration of 300 nM (data not shown). However, none of the primers sets were able to amplify genomic DNA isolated from mouse tails (using the method described in Chapter IV Materials and Methods). This protocol for isolation of DNA from tails was appropriate for other PCR reactions, as demonstrated in Figure 20. DNA from control and transgenic animals was then purified with a standard phenol-choloform extraction, precipitated with ethanol, and resuspended in TE buffer. This purified DNA was tested with the real-time primers in the Elsea lab ABI thermocycler and amplification was noted, though not consistently in all samples. BAC 326M22 DNA used as a control in these reactions was amplified. When 50 or 100 ng of purified mouse genomic DNA were tested under real-time cycling conditions in the ABI 7700, there was either no amplification above baseline or all of the replicates produced different C_T values. There was no consistent amplification that could be analyzed. Possible reasons for this failure are: the original concentration of the genomic DNA was not correct, a contaminant such as salt from the DNA extraction was inhibiting the PCR reaction, or perhaps a great deal of error was created within the replicate samples when diluting the DNA. The genomic DNA was tested twice with the Sp6 BAC end primers and while the

BAC DNA controls were amplified, the genomic DNA samples failed. Perhaps a different method of isolating genomic DNA from the mice would provide better results. We have had variable success in the Elsea laboratory isolating usable DNA from mouse whole blood, probably due to blood collection tubes containing heparin, which can inhibit several of the standard molecular procedures that are routinely performed in the lab such as PCR and restriction digests. In the future, if a viable blood collection method is devised without heparin, genomic DNA extracted from mouse blood may provide better real-time PCR amplification and resolve the integrated *Rai1* BAC copy number.

BIBLIOGRAPHY

Aasland, R., Gibson, T. J. and Stewart, A. F. (1995). "The PHD finger: implications for chromatin-mediated transcriptional regulation." Trends Biochem Sci 20(2): 56-9.

Ackerman, S. H. (2002). "Atp11p and Atp12p are chaperones for F(1)-ATPase biogenesis in mitochondria." Biochim Biophys Acta 1555(1-3): 101-5.

Alagille, D., Odievre, M., Gautier, M. and Dommergues, J. P. (1975). "Hepatic ductular hypoplasia associated with characteristic facies, vertebral malformations, retarded physical, mental, and sexual development, and cardiac murmur." J Pediatr 86(1): 63-71.

Alavizadeh, A., Kiernan, A. E., et al. (2001). "The Wheels mutation in the mouse causes vascular, hindbrain, and inner ear defects." Dev Biol 234(1): 244-60.

Bi, W., Park, S. S., et al. (2003). "Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2." Am J Hum Genet 73(6): 1302-15.

Bi, W., Yan, J., et al. (2002). "Genes in a refined Smith-Magenis syndrome critical deletion interval on chromosome 17p11.2 and the syntenic region of the mouse." Genome Res 12(5): 713-28.

Butler, M. G. (2002). "Imprinting disorders: non-Mendelian mechanisms affecting growth." J Pediatr Endocrinol Metab 15 Suppl 5: 1279-88.

Campbell, H. D., Fountain, S., et al. (2002). "Fliih, a gelsolin-related cytoskeletal regulator essential for early mammalian embryonic development." Mol Cell Biol 22(10): 3518-26.

Campbell, H. D., Fountain, S., et al. (1997). "Genomic structure, evolution, and expression of human FLII, a gelsolin and leucine-rich-repeat family member: overlap with LLGL." Genomics 42(1): 46-54.

Campbell, H. D., Fountain, S., et al. (2000). "Fliih, the murine homologue of the Drosophila melanogaster flightless I gene: nucleotide sequence, chromosomal mapping and overlap with Llglh." DNA Seq 11(1-2): 29-40.

Cassidy, S. B., Dykens, E. and Williams, C. A. (2000). "Prader-Willi and Angelman syndromes: sister imprinted disorders." Am J Med Genet 97(2): 136-46.

Chen, K. S., Gunaratne, P. H., et al. (1995). "The human homologue of the Drosophila melanogaster flightless-I gene (flil) maps within the Smith-Magenis microdeletion critical region in 17p11.2." Am J Hum Genet 56(1): 175-82.

Chen, K. S., Manian, P., et al. (1997). "Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome." Nat Genet 17(2): 154-63.

Cismowski, M. J., Ma, C., et al. (2000). "Activation of heterotrimeric G-protein signaling by a ras-related protein. Implications for signal integration." J Biol Chem 275(31): 23421-4.

Cismowski, M. J., Takesono, A., et al. (1999). "Genetic screens in yeast to identify mammalian nonreceptor modulators of G-protein signaling." Nat Biotechnol 17(9): 878-83.

Crawley, J. N. (2000). What's wrong with my mouse?: Behavioral phenotyping of transgenic and knockout mice, John Wiley and Sons, Inc.

De Leersnyder, H., De Blois, M. C., et al. (2001). "Inversion of the circadian rhythm of melatonin in the Smith-Magenis syndrome." J Pediatr 139(1): 111-6.

De Leersnyder, H., de Blois, M. C., et al. (2001). "beta(1)-adrenergic antagonists improve sleep and behavioural disturbances in a circadian disorder, Smith-Magenis syndrome." J Med Genet 38(9): 586-90.

DeSilva, U., Massa, H., Trask, B. J. and Green, E. D. (1999). "Comparative mapping of the region of human chromosome 7 deleted in williams syndrome." Genome Res 9(5): 428-36.

Elsea, S. H., Fritz, E., Schoener-Scott, R., Meyn, M. S. and Patel, P. I. (1998). "Gene for topoisomerase III maps within the Smith-Magenis syndrome critical region: analysis of cell-cycle distribution and radiation sensitivity." Am J Med Genet 75(1): 104-8.

Elsea, S. H., Juyal, R. C., et al. (1995). "Haploinsufficiency of cytosolic serine hydroxymethyltransferase in the Smith-Magenis syndrome." Am J Hum Genet 57(6): 1342-50.

Elsea, S. H. and Lucas, R. E. (2002). "The mousetrap: what we can learn when the mouse model does not mimic the human disease." ILAR J 43(2): 66-79.

Elsea, S. H., Mykytyn, K., et al. (1999). "Hemizygosity for the COP9 signalosome subunit gene, SGN3, in the Smith- Magenis syndrome." Am J Med Genet 87(4): 342-8.

Elsea, S. H., Purandare, S. M., et al. (1997). "Definition of the critical interval for Smith-Magenis syndrome." Cytogenet Cell Genet 79(3-4): 276-81.

Fang, M., Jaffrey, S. R., et al. (2000). "Dexras1: a G protein specifically coupled to neuronal nitric oxide synthase via CAPON." Neuron 28(1): 183-93.

Farrell, M. J., Stadt, H., et al. (1999). "HIRA, a DiGeorge syndrome candidate gene, is required for cardiac outflow tract septation." Circ Res 84(2): 127-35.

Finucane, B., Dirrigl, K. H. and Simon, E. W. (2001). "Characterization of self-injurious behaviors in children and adults with Smith-Magenis syndrome." Am J Ment Retard 106(1): 52-8.

Francke, U. (1999). "Williams-Beuren syndrome: genes and mechanisms." Hum Mol Genet 8(10): 1947-54.

Franke, Y., Peoples, R. J. and Francke, U. (1999). "Identification of GTF2IRD1, a putative transcription factor within the Williams-Beuren syndrome deletion at 7q11.23." Cytogenet Cell Genet 86(3-4): 296-304.

Fritz, E., Elsea, S. H., Patel, P. I. and Meyn, M. S. (1997). "Overexpression of a truncated human topoisomerase III partially corrects multiple aspects of the ataxia-telangiectasia phenotype." Proc Natl Acad Sci U S A 94(9): 4538-42.

Funke, B., Epstein, J. A., et al. (2001). "Mice overexpressing genes from the 22q11 region deleted in velo-cardio-facial syndrome/DiGeorge syndrome have middle and inner ear defects." Hum Mol Genet 10(22): 2549-56.

Gilbert, S. (2000). Developmental Biology, Sinauer Associates, Inc.

Greenberg, F., Guzzetta, V., et al. (1991). "Molecular analysis of the Smith-Magenis syndrome: a possible contiguous- gene syndrome associated with del(17)(p11.2)." Am J Hum Genet 49(6): 1207-18.

Greenberg, F., Lewis, R. A., et al. (1996). "Multi-disciplinary clinical study of Smith-Magenis syndrome (deletion 17p11.2)." Am J Med Genet 62(3): 247-54.

Gridley, T. (2003). "Notch signaling and inherited disease syndromes." Hum Mol Genet 12 Spec No 1: R9-13.

Guris, D. L., Fantes, J., Tara, D., Druker, B. J. and Imamoto, A. (2001). "Mice lacking the homologue of the human 22q11.2 gene CRKL phenocopy neurocristopathies of DiGeorge syndrome." Nat Genet 27(3): 293-8.

Guzzetta, V., Franco, B., et al. (1992). "Somatic cell hybrids, sequence-tagged sites, simple repeat polymorphisms, and yeast artificial chromosomes for physical and genetic mapping of proximal 17p." Genomics 13(3): 551-9.

Hanai, R., Caron, P. R. and Wang, J. C. (1996). "Human TOP3: a single-copy gene encoding DNA topoisomerase III." Proc Natl Acad Sci U S A 93(8): 3653-7.

Hayes, S., Turecki, G., et al. (2000). "CAG repeat length in RAI1 is associated with age at onset variability in spinocerebellar ataxia type 2 (SCA2)." Hum Mol Genet 9(12): 1753-8.

Heintz, N. (2001). "BAC to the future: the use of bac transgenic mice for neuroscience research." Nat Rev Neurosci 2(12): 861-70.

Hockenhull, E. L., Carette, M. J., et al. (1999). "A complete physical contig and partial transcript map of the Williams syndrome critical region." Genomics 58(2): 138-45.

Hoogenraad, C. C., Eussen, B. H., et al. (1998). "The murine CYLN2 gene: genomic organization, chromosome localization; and comparison to the human gene that is located within the 7q11.23 Williams syndrome critical region." Genomics 53(3): 348-58.

Hoogenraad, C. C., Koekkoek, B., et al. (2002). "Targeted mutation of Cyln2 in the Williams syndrome critical region links CLIP-115 haploinsufficiency to neurodevelopmental abnormalities in mice." Nat Genet 32(1): 116-27.

Imai, Y., Suzuki, Y., et al. (1995). "Cloning of a retinoic acid-induced gene, GT1, in the embryonal carcinoma cell line P19: neuron-specific expression in the mouse brain." Brain Res Mol Brain Res 31(1-2): 1-9.

Ishikawa, K., Azuma, S., et al. (2003). "Cloning and characterization of Xenopus laevis drg2, a member of the developmentally regulated GTP-binding protein subfamily." Gene 322: 105-12.

Jerome, L. A. and Papaioannou, V. E. (2001). "DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1." Nat Genet 27(3): 286-91.

Ji, Y., Eichler, E. E., Schwartz, S. and Nicholls, R. D. (2000). "Structure of chromosomal duplicons and their role in mediating human genomic disorders." Genome Res 10(5): 597-610.

Joober, R., Benkelfat, C., et al. (1999). "Analysis of 14 CAG repeat-containing genes in schizophrenia." Am J Med Genet 88(6): 694-9.

Juyal, R. C., Figuera, L. E., et al. (1996). "Molecular analyses of 17p11.2 deletions in 62 Smith-Magenis syndrome patients." Am J Hum Genet 58(5): 998-1007.

Kallioniemi, O. P., Kallioniemi, A., et al. (1994). "Physical mapping of chromosome 17 cosmids by fluorescence in situ hybridization and digital image analysis." Genomics 20(1): 125-8.

Kao, F., Jones, C. and Puck, T. T. (1976). "Genetics of somatic mammalian cells: genetic, immunologic, and biochemical analysis with Chinese hamster cell hybrids containing selected human chromosomes." PNAS 73: 193-97.

Kapfhamer, D., Valladares, O., et al. (2002). "Mutations in Rab3a alter circadian period and homeostatic response to sleep loss in the mouse." Nat Genet 32(2): 290-5.

Karmiloff-Smith, A., Grant, J., et al. (2003). "Using case study comparisons to explore genotype-phenotype correlations in Williams-Beuren syndrome." J Med Genet 40(2): 136-40.

Kemppainen, R. J. and Behrend, E. N. (1998). "Dexamethasone rapidly induces a novel ras superfamily member-related gene in AtT-20 cells." J Biol Chem 273(6): 3129-31.

Kemppainen, R. J., Cox, E., Behrend, E. N., Brogan, M. D. and Ammons, J. M. (2003). "Identification of a glucocorticoid response element in the 3'-flanking region of the human Dexras1 gene." Biochim Biophys Acta 1627(2-3): 85-9.

Khan, N. L. and Wood, N. W. (1999). "Prader-Willi and Angelman syndromes: update on genetic mechanisms and diagnostic complexities." Curr Opin Neurol 12(2): 149-54.

Kiernan, A. E., Ahituv, N., et al. (2001). "The Notch ligand Jagged1 is required for inner ear sensory development." Proc Natl Acad Sci U S A 98(7): 3873-8.

Kishino, T., Lalande, M. and Wagstaff, J. (1997). "UBE3A/E6-AP mutations cause Angelman syndrome." Nat Genet 15(1): 70-3.

Krantz, I. D., Piccoli, D. A. and Spinner, N. B. (1997). "Alagille syndrome." J Med Genet 34(2): 152-7.

Krantz, I. D., Smith, R., et al. (1999). "Jagged1 mutations in patients ascertained with isolated congenital heart defects." Am J Med Genet 84(1): 56-60.

Kuklin, A., Davis, A. P., Hecker, K. H., Gjerde, D. T. and Taylor, P. D. (1999). "A novel technique for rapid automated genotyping of DNA polymorphisms in the mouse." Mol Cell Probes 13(3): 239-42.

Kuklin, A., Munson, K., Gjerde, D., Haefele, R. and Taylor, P. (1997). "Detection of single-nucleotide polymorphisms with the WAVE DNA fragment analysis system." Genet Test 1(3): 201-6.

Lamour, V., Lecluse, Y., et al. (1995). "A human homolog of the S. cerevisiae HIR1 and HIR2 transcriptional repressors cloned from the DiGeorge syndrome critical region." Hum Mol Genet 4(5): 791-9.

Le Caignec, C., Lefevre, M., et al. (2002). "Familial deafness, congenital heart defects, and posterior embryotoxon caused by cysteine substitution in the first epidermal-growth-factor-like domain of jagged 1." Am J Hum Genet 71(1): 180-6.

Lee, S. H., Fu, K. K., Hui, J. N. and Richman, J. M. (2001). "Noggin and retinoic acid transform the identity of avian facial prominences." Nature 414(6866): 909-12.

Lemckert, F. A., Sedgwick, J. D. and Korner, H. (1997). "Gene targeting in C57BL/6 ES cells. Successful germ line transmission using recipient BALB/c blastocysts developmentally matured in vitro." Nucleic Acids Res 25(4): 917-8.

Li, D. Y., Brooke, B., et al. (1998). "Elastin is an essential determinant of arterial morphogenesis." Nature 393(6682): 276-80.

Li, D. Y., Faury, G., et al. (1998). "Novel arterial pathology in mice and humans hemizygous for elastin." J Clin Invest 102(10): 1783-7.

Li, L., Krantz, I. D., et al. (1997). "Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1 [see comments]." Nat Genet 16(3): 243-51.

Li, Q., Liu, Z., Monroe, H. and Culiat, C. T. (2002). "Integrated platform for detection of DNA sequence variants using capillary array electrophoresis." Electrophoresis 23(10): 1499-511.

Liang, S. G., Sadovnick, A. D., et al. (2002). "A linkage disequilibrium study of bipolar disorder and microsatellite markers on 22q13." Psychiatr Genet 12(4): 231-5.

Liang, Y., Wang, A., et al. (1999). "Characterization of the Human and Mouse Unconventional Myosin XV Genes Responsible for Hereditary Deafness DFNB3 and Shaker 2." Genomics 61(3): 243-58.

Liang, Y., Wang, A., et al. (1999). "Characterization of the human and mouse unconventional myosin XV genes responsible for hereditary deafness DFNB3 and shaker 2." Genomics 61(3): 243-58.

Liburd, N., Ghosh, M., et al. (2001). "Novel mutations of MYO15A associated with profound deafness in consanguineous families and moderately severe hearing loss in a patient with Smith-Magenis syndrome." Hum Genet 109(5): 535-41.

Lindsay, E. A. (2001). "Chromosomal microdeletions: dissecting del22q11 syndrome." Nat Rev Genet 2(11): 858-68.

Lindsay, E. A. and Baldini, A. (1998). "Congenital heart defects and 22q11 deletions: which genes count?" Mol Med Today 4(8): 350-7.

Lindsay, E. A., Vitelli, F., et al. (2001). "Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice." Nature 410(6824): 97-101.

Lohi, O. and Lehto, V. P. (1998). "VHS domain marks a group of proteins involved in endocytosis and vesicular trafficking." FEBS Lett 440(3): 255-7.

Lohi, O., Poussu, A., Mao, Y., Quiocho, F. and Lehto, V. P. (2002). "VHS domain -- a longshoreman of vesicle lines." FEBS Lett 513(1): 19-23.

Lucas, R. E., Vlangos, C. N., Das, P., Patel, P. I. and Elsea, S. H. (2001). "Genomic organisation of the approximately 1.5 Mb Smith-Magenis syndrome critical interval: transcription map, genomic contig, and candidate gene analysis." Eur J Hum Genet 9(12): 892-902.

Lyngso, C., Bouteiller, G., et al. (2000). "Interaction between the transcription factor SPBP and the positive cofactor RNF4. An interplay between protein binding zinc fingers." J Biol Chem 275(34): 26144-9.

Maden, M. (2002). "Retinoid signalling in the development of the central nervous system." Nat Rev Neurosci 3: 843-53.

Maden, M. and Hind, M. (2003). "Retinoic acid, a regeneration-inducing molecule." Dev Dyn 226(2): 237-44.

Mann, M. R. and Bartolomei, M. S. (1999). "Towards a molecular understanding of Prader-Willi and Angelman syndromes." Hum Mol Genet 8(10): 1867-73.

Matsuura, T., Sutcliffe, J. S., et al. (1997). "De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome." Nat Genet 15(1): 74-7.

McCright, B., Gao, X., et al. (2001). "Defects in development of the kidney, heart and eye vasculature in mice homozygous for a hypomorphic Notch2 mutation." Development 128(4): 491-502.

Merscher, S., Funke, B., et al. (2001). "TBX1 is responsible for cardiovascular defects in velo-cardio- facial/DiGeorge syndrome." Cell 104(4): 619-29.

Nolan, P. M., Kapfhamer, D. and Bucan, M. (1997). "Random mutagenesis screen for dominant behavioral mutations in mice." Methods 13(4): 379-95.

Oda, T., Elkahloun, A. G., Meltzer, P. S. and Chandrasekharappa, S. C. (1997). "Identification and cloning of the human homolog (JAG1) of the rat Jagged1 gene from the Alagille syndrome critical region at 20p12." Genomics 43(3): 376-9.

Oda, T., Elkahloun, A. G., et al. (1997). "Mutations in the human Jagged1 gene are responsible for Alagille syndrome [see comments]." Nat Genet 16(3): 235-42.

Ogdie, M. N., Macphie, I. L., et al. (2003). "A genomewide scan for attention-deficit/hyperactivity disorder in an extended sample: suggestive linkage on 17p11." Am J Hum Genet 72(5): 1268-79.

Ohta, T., Buiting, K., et al. (1999). "Molecular mechanism of angelman syndrome in two large families involves an imprinting mutation." Am J Hum Genet 64(2): 385-96.

Osborne, L. R. (1999). "Williams-Beuren syndrome: unraveling the mysteries of a microdeletion disorder." Mol Genet Metab 67(1): 1-10.

Paglia, D. E., Valentine, W. N. and Brockway, R. A. (1984). "Identification of thymidine nucleotidase and deoxyribonucleotidase activities among normal isozymes of 5'-nucleotidase in human erythrocytes." Proc Natl Acad Sci U S A 81(2): 588-92.

Park, S. S., Stankiewicz, P., et al. (2002). "Structure and evolution of the Smith-Magenis syndrome repeat gene clusters, SMS-REPs." Genome Res 12(5): 729-38.

Pascual, J., Martinez-Yamout, M., Dyson, H. J. and Wright, P. E. (2000). "Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor." J Mol Biol 304(5): 723-9.

Paylor, R., McIlwain, K. L., et al. (2001). "Mice deleted for the DiGeorge/velocardiofacial syndrome region show abnormal sensorimotor gating and learning and memory impairments." Hum Mol Genet 10(23): 2645-50.

Peoples, R., Franke, Y., et al. (2000). "A physical map, including a BAC/PAC clone contig, of the Williams- Beuren syndrome--deletion region at 7q11.23." Am J Hum Genet 66(1): 47-68.

Peoples, R. J., Cisco, M. J., Kaplan, P. and Francke, U. (1998). "Identification of the WBSCR9 gene, encoding a novel transcriptional regulator, in the Williams-Beuren syndrome deletion at 7q11.23." Cytogenet Cell Genet 82(3-4): 238-46.

Perez Jurado, L. A., Peoples, R., Kaplan, P., Hamel, B. C. and Francke, U. (1996). "Molecular definition of the chromosome 7 deletion in Williams syndrome and parent-of-origin effects on growth." Am J Hum Genet 59(4): 781-92.

Petrij, F., Giles, R. H., et al. (1995). "Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP." Nature 376(6538): 348-51.

Piccoli, D. A. and Spinner, N. B. (2001). "Alagille syndrome and the Jagged1 gene." Semin Liver Dis 21(4): 525-34.

Pickard, G. E., Sollars, P. J., Rinchik, E. M., Nolan, P. M. and Bucan, M. (1995). "Mutagenesis and behavioral screening for altered circadian activity identifies the mouse mutant, Wheels." Brain Res 705(1-2): 255-66.

Pickova, A., Paul, J., Petruzzella, V. and Houstek, J. (2003). "Differential expression of ATPAF1 and ATPAF2 genes encoding F(1)-ATPase assembly proteins in mouse tissues." FEBS Lett 551(1-3): 42-6.

Pizzuti, A., Novelli, G., et al. (1997). "UFD1L, a developmentally expressed ubiquitination gene, is deleted in CATCH 22 syndrome." Hum Mol Genet 6(2): 259-65.

Pollet, N., Dhorne-Pollet, S., et al. (1995). "Construction of a 3.7-Mb physical map within human chromosome 20p12 ordering 18 markers in the Alagille syndrome locus." Genomics 27(3): 467-74.

Potocki, L., Chen, K. S. and Lupski, J. R. (1999). "Subunit 3 of the COP9 signal transduction complex is conserved from plants to humans and maps within the smith-magenis syndrome critical region in 17p11.2." Genomics 57(1): 180-2.

Potocki, L., Chen, K. S., et al. (2000). "Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion." Nat Genet 24(1): 84-7.

Potocki, L., Glaze, D., et al. (2000). "Circadian rhythm abnormalities of melatonin in Smith-Magenis syndrome." J Med Genet 37(6): 428-33.

Probst, F. J., Fridell, R. A., et al. (1998). "Correction of deafness in shaker-2 mice by an unconventional myosin in a BAC transgene." Science 280(5368): 1444-7.

Rampazzo, C., Gallinaro, L., et al. (2000). "A deoxyribonucleotidase in mitochondria: involvement in regulation of dNTP pools and possible link to genetic disease." Proc Natl Acad Sci U S A 97(15): 8239-44.

Rekdal, C., Sjottem, E. and Johansen, T. (2000). "The nuclear factor SPBP contains different functional domains and stimulates the activity of various transcriptional activators." J Biol Chem 275(51): 40288-300.

Roberts, C., Daw, S. C., Halford, S. and Scambler, P. J. (1997). "Cloning and developmental expression analysis of chick Hira (Chira), a candidate gene for DiGeorge syndrome." Hum Mol Genet 6(2): 237-45.

Rougeulle, C., Glatt, H. and Lalande, M. (1997). "The Angelman syndrome candidate gene, UBE3A/E6-AP, is imprinted in brain." Nat Genet 17(1): 14-5.

Sazuka, T., Kinoshita, M., et al. (1992). "Expression of DRG during murine embryonic development." Biochem Biophys Res Commun 189(1): 371-7.

Sazuka, T., Tomooka, Y., Ikawa, Y., Noda, M. and Kumar, S. (1992). "DRG: a novel developmentally regulated GTP-binding protein." Biochem Biophys Res Commun 189(1): 363-70.

Schedl, A., Larin, Z., et al. (1993). "A method for the generation of YAC transgenic mice by pronuclear microinjection." Nucleic Acids Res 21(20): 4783-7.

Schenker, T., Lach, C., Kessler, B., Calderara, S. and Trueb, B. (1994). "A novel GTP-binding protein which is selectively repressed in SV40 transformed fibroblasts." J Biol Chem 269(41): 25447-53.

Schmickel, R. D. (1986). "Contiguous gene syndromes: a component of recognizable syndromes." J Pediatr 109: 231-41.

Schuster-Gossler, K., Lee, A. W., et al. (2001). "Use of coisogenic host blastocysts for efficient establishment of germline chimeras with C57BL/6J ES cell lines." Biotechniques 31(5): 1022-4, 26.

Seranski, P., Hoff, C., et al. (2001). "RAI1 is a novel polyglutamine encoding gene that is deleted in Smith-Magenis syndrome patients." Gene 270(1-2): 69-76.

Shapira, S. K. (1998). "An update on chromosome deletion and microdeletion syndromes." Curr Opin Pediatr 10(6): 622-7.

Shiba, Y., Katoh, Y., et al. (2003). "GAT (GGA and Tom1) domain responsible for ubiquitin binding and ubiquitination." J Biol Chem.

Shimano, H., Shimomura, I., et al. (1997). "Elevated levels of SREBP-2 and cholesterol synthesis in livers of mice homozygous for a targeted disruption of the SREBP-1 gene." J Clin Invest 100(8): 2115-24.

Slager, R. E., Newton, T. L., Vlangos, C. N., Finucane, B. and Elsea, S. H. (2003). "Mutations in RAI1 associated with Smith-Magenis syndrome." Nat Genet 33(4): 466-8.

Smith, A. C., Dykens, E. and Greenberg, F. (1998). "Behavioral phenotype of Smith-Magenis syndrome (del 17p11.2)." Am J Med Genet 81(2): 179-85.

Smith, A. C., McGavran, L., et al. (1986). "Interstitial deletion of (17)(p11.2p11.2) in nine patients." Am J Med Genet 24(3): 393-414.

Smith, A. C. M., L., M., Waldstein, G. and Robinson, J. (1982). "Deletion of the 17 short arm in two patients with facial clefts and congenital heart disease." Am J Hum Genet 34: A410.

Smith, J. R., Osborne, T. F., Goldstein, J. L. and Brown, M. S. (1990). "Identification of nucleotides responsible for enhancer activity of sterol regulatory element in low density lipoprotein receptor gene." J Biol Chem 265(4): 2306-10.

Sonuga-Barke, E. J. (2003). "The dual pathway model of AD/HD: an elaboration of neuro-developmental characteristics." Neurosci Biobehav Rev 27(7): 593-604.

Spinner, N. B., Colliton, R. P., et al. (2001). "Jagged1 mutations in alagille syndrome." Hum Mutat 17(1): 18-33.

Stankiewicz, P. and Lupski, J. R. (2002). "Molecular-evolutionary mechanisms for genomic disorders." Curr Opin Genet Dev 12(3): 312-9.

Stankiewicz, P., Shaw, C. J., et al. (2003). "Genome architecture catalyzes nonrecurrent chromosomal rearrangements." Am J Hum Genet 72(5): 1101-16.

Strand, D., Unger, S., et al. (1995). "A human homologue of the Drosophila tumour suppressor gene l(2)gl maps to 17p11.2-12 and codes for a cytoskeletal protein that associates with nonmuscle myosin II heavy chain." Oncogene 11(2): 291-301.

Stratton, R. F., Dobyns, W. B., et al. (1986). "Interstitial deletion of (17)(p11.2p11.2): report of six additional patients with a new chromosome deletion syndrome." Am J Med Genet 24(3): 421-32.

Surks, H. K., Richards, C. T. and Mendelsohn, M. E. (2003). "Myosin phosphatase-Rho interacting protein. A new member of the myosin phosphatase complex that directly binds RhoA." J Biol Chem 278(51): 51484-93.

Taddei, I., Morishima, M., Huynh, T. and Lindsay, E. A. (2001). "Genetic factors are major determinants of phenotypic variability in a mouse model of the DiGeorge/del22q11 syndromes." Proc Natl Acad Sci U S A 98(20): 11428-31.

Takahashi, H., Umeda, N., et al. (2003). "Mouse dexamethasone-induced RAS protein 1 gene is expressed in a circadian rhythmic manner in the suprachiasmatic nucleus." Brain Res Mol Brain Res 110(1): 1-6.