GENETIC AND PHYSICAL MAPPING OF THE GENE FOR HALLERVORDEN-SPATZ SYNDROME

by

Todd Duane Taylor

A DISSERTATION

Presented to the Department of Molecular and Medical Genetics

and the Oregon Health Sciences University

School of Medicine

Portland, Oregon

in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

December 1998

School of Medicine

Oregon Health Sciences University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. thesis of

Todd Duane Taylor

has been approved

[redacted]

Professor in charge of thesis

[redacted]

Member

[redacted]

Member

[redacted]

Member

[redacted]

Member

[redacted]

Associate Dean for Graduate Studies

TABLE OF CONTENTS

TABLES AND ILLUSTRATIONS

Appendices

ABBREVIATIONS

ATP	adenosine 5'-triphosphate
BAC	bacterial artificial chromosome
bp	base pair
cDNA	complementary deoxyribonucleic acid
CEPH	Centre d'Etude du Polymorphisme Humain
CHLC	Cooperative Human Linkage Center
cM	centimorgans
CNS	central nervous system
CS	Clouston syndrome
CSF	cerebrospinal fluid
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
EtBr	ethidium bromide
GABA	gamma aminobutyric acid
GABAT	gamma aminobutyric acid transaminase
GAD	glutamic acid decarboxylase
GDB	Genome Database
GP	globus pallidus
HSS	Hallervorden-Spatz syndrome
kb	kilobase (1000 DNA bases)
LB	Luria-Bertani medium
MAG	myelin-associated glycoprotein
Mb	megabase (1,000,000 DNA bases)

MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
NBIA1	Neurodegeneration with Brain Iron Accumulation, Type 1
NCBI	National Center for Biotechnology Information
NIH	National Institutes of Health
NIRCA	non-isotopic RNase cleavage assay
nt	nucleotide
OHSU	Oregon Health Sciences University
oligo	oligonucleotide
OMIM	Online Mendelian Inheritance in Man
PAC	P1 artificial chromosome
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
RH	radiation hybrid
RNA	ribonucleic acid
RNase	ribonuclease
RP	retinitis pigmentosa
RT-PCR	reverse transcription/polymerase chain reaction
SHGC	Stanford Human Genome Center
SNR	substantia nigra reticular zone
STS	sequence-tagged site
TBE	Tris/borate EDTA (buffer; 0.089 M Tris, 0.089 Borate, and 0.002 M EDTA)
TIGR	The Institute for Genomic Research
TRE	triplet repeat expansion
Tris	tris(hydroxymethyl)aminomethane
Tris-Cl	Tris hydrochloride
UCSF	University of California, San Francisco

VNTR	variable number of tandem repeats polymorphism
WICGR	Whitehead Institute/MIT Center for Genome Research
YAC	yeast artificial chromosome

ACKNOWLEDGEMENTS

I am grateful to all the people who made this work possible. Foremost, I would like to thank my mentor, Dr. Susan Hayflick, who was daring enough to accept me as her first graduate student. She was always very enthusiastic about the project and the results, whether positive or negative. She encouraged me to learn as much as possible and was always available when I needed her help or advice (something I now realize I should have taken more advantage of). Without her tolerance and support, this work not have come as far as it did. I can never repay her for the opportunity she gave me.

I would like to thank Dr. Mike Litt for giving me a lab to call home, and for all of his input on all aspects of this project. The resources he provided were invaluable and greatly appreciated. Dr. Litt's vast knowledge of genetics and laboratory techniques make him and endless source for information.

Thank you to Dr. Patti Kramer, who was kind enough to teach me linkage analysis and to provide the staff and technical support required to run it. She was always there when we needed an unbiased eye to examine our data. She was also very encouraging and supportive.

Thank you to Dr. Jon Zonana for asking me to be involved with the Clouston syndrome project. When the work with HSS was going somewhat slow, this side project kept me busy.

Thank you to the many technicians (Eric Smith, Thas Phromchotikul, Dante La Morticella) and post-doctorates (David Browne, Roque Carrero, Dennis Schultz) in Dr.

Litt's lab for their advice and assistance.

A special thank you to Amy Malone, a technician working for Dr. Hayflick, who came along when things were getting way too busy for me to handle alone.

Thank you to the members of my thesis and qualifying committees for all their guidance and recommendations in this project.

Thank you to the Molecular and Medical Genetics department for taking a chance on an old math major and for letting me do my graduate work at OHSU. Thanks to all the staff for their support and help along the way.

I would like to acknowledge our collaborators, Jane Gitschier, Bing Zhou, and Barbara Levinson at the University of California San Francisco. They have taken on several parts of the project that we would have been overwhelmed with.

I would especially like to thank my wife, Miyoko, for supporting me throughout this entire process and for giving me a son, Kianu, who has provided me with that extra bit of inspiration when I really needed it. Also, thanks to both my Japanese (Nobuko Tanaka) and American (Janet and Duane Taylor, Teri and Scott Deshler) families for their unending support.

Lastly, thank you to the Big Kahuna and to anyone else that I may have forgotten that lent support, knowledge, or encouragement along the way. Domo arigato gozaimashita!

ABSTRACT

Hallervorden-Spatz syndrome (HSS, OMIM #234200) is a rare, autosomal recessive disorder with brain iron accumulation as a prominent finding. It is characterized by rigidity, dystonia, pigmentary retinopathy, and basal ganglia densities on magnetic resonance imaging. Onset is in childhood with a progressive course leading to death by early adulthood. The objectives of this study were to determine the chromosomal location of the HSS gene using linkage analysis and to characterize the genomic region containing the HSS gene. Samples were initially obtained from 10 families, one each of Amish, New Zealand and Australian descent, two of Spanish descent, and five of Italian descent. Highly polymorphic microsatellite markers near candidate genes encoding proteins related to iron homeostasis and retinitis pigmentosa were typed first, and these regions were excluded. A primary genome search for linkage to the HSS gene was performed using samples from the large, consanguineous, multi-generation Amish family. Two-point linkage analysis was run using MLINK and haplotypes were constructed for all markers. Linkage was demonstrated between the Amish family and genetic markers from chromosome 20p13-p12.3. Analysis of the other nine families supported linkage of the HSS gene to this region with a total maximum two-point lod score of 13.75 at $\theta = 0$ for the marker AFMa049yd1. Homozygosity in the Amish family and obligate recombinants in three of the other families narrowed the HSS gene to a 4 centimorgan interval between markers D20S906 and D20S116. This locus was designated Neurodegeneration with Brain Iron Accumulation, Type 1 (NBIA1). Correlation between the publicly available genetic, radiation hybrid and yeast artificial chromosome (YAC) maps from this region was very poor. Very few markers were on all of the maps and none of the maps was

sufficiently marker dense. A high-resolution radiation hybrid panel was first used to integrate markers from these maps. A minimal set of YACs spanning the NBIA1 critical region was available and was typed for all markers. Several of the YACs were found to have internal deletions. To confidently map all of the markers in the NBIA1 critical region, a human bacterial artificial chromosome (BAC) contig was developed. BAC clones are much more stable and less chimeric than YAC clones. A preliminary BAC/YAC contig across the HSS critical region was constructed, spanning more than 3 megabases. Several genes and uncharacterized expressed sequence tags previously mapped to this region were excluded from the HSS candidate region. An additional 50 families with classical and non-classical HSS were identified and those with the classical type were haplotyped. Newly identified obligate recombinants, regions of homozygosity, and shared haplotypes suggestive of linkage disequilibrium, further delineated the HSS interval to that between markers D20S473 and D20S867. This interval is spanned by two overlapping BACs with a total size of 400 kilobases (kb), which suggests that the current Hallervorden-Spatz syndrome critical region is less than 400 kb.

CHAPTER 1

INTRODUCTION

OVERVIEW

Hallervorden-Spatz syndrome (HSS, OMIM #234200) is a rare, autosomal recessive disorder with brain iron accumulation as a prominent finding. It was first described in 1922 by Julius Hallervorden and Hugo Spatz. HSS is characterized by rigidity, dystonia, and pigmentary retinopathy. The onset in childhood with a progressive course leading to death by early adulthood. Little is known about the biochemical or cellular defects in this disease. Iron accumulates intra- and extracellulary in the basal ganglia, specifically in the globus pallidus and substantia nigra pars reticulata. Peripheral iron homeostasis is normal, as are cerebrospinal fluid (CSF) levels of iron, transferrin and ferritin. Axonal spheroid formations are found on pathological examination of affected areas of brain. Acanthocytes and lipofuscin-laden histiocytes are seen in a subset of patients with HSS.

To elucidate the defect in HSS, we took a genetic approach using homozygosity mapping (Lander and Botstein, 1987). This method has been used to successfully map genes for other rare disorders (Farrall, 1993) including Bardet-Biedl syndrome (Kwitek-Black et al., 1993), familial Mediterranean fever (Aksentijevich et al., 1993), Friedreich's ataxia phenotype with selective vitamin E deficiency (Ben Hamida et al., 1993), and Hirschsprung disease (Puffenberger et al., 1994). To reduce the possibility of locus heterogeneity, linkage analysis to the HSS gene was performed using samples from just one family. This family, HS1 (**Figure 1**), is a large, inbred Amish family with multiple consanguineous unions. The HS1 family alone, due to its size and degree of inbreeding, provides sufficient statistical power to establish linkage to one chromosomal region. Highly polymorphic microsatellite markers were typed near candidate genes encoding proteins of iron homeostasis and pigmentary retinopathy, and subsequently across the

whole autosomal genome. Two-point linkage analysis and haplotype analysis were used to exclude or confirm linkage. Other methods, such as multipoint linkage analysis and linkage disequilibrium analysis were also considered in the localization of the HSS gene. Additional classical HSS families were also genotyped and incorporated into the analyses.

Once linkage was established to chromosome 20, it was necessary to construct a physical map across the HSS gene critical region. The correlation between numerous available maps of chromosome 20 was poor; only a few markers were constantly ordered and on all of the maps. By improving the mapping information across the HSS critical region, we expected to rectify marker order and to consolidate all of the marker information from the publicly available maps. A high-resolution radiation hybrid panel was first used to integrate markers from these maps. YACs spanning the critical region were also typed with the same markers. Since several of the YACs were found to be internally deleted, we decided to screen a BAC DNA library in order to build a more reliable contig. BACs are much more stable and less chimeric than YAC clones. A contig spanning more than 3 megabases was constructed across the HSS critical region. We were able to exclude most of the genes previously mapped to the region from the HSS critical region with this contig. After genotyping several new classical HSS families and improving the physical map, we were able to narrow the HSS critical region down to two overlapping BACs with a total size of 400 kb.

Before World War II, Julius Hallervorden was shamelessly and actively involved in the euthanasia of 'mental defectives', whose brains he collected for study. For this reason, we and others (Shevell, 1992; Gordon, 1993) proposed that any genes for this disorder

not carry his name. In compliance, this first locus for Hallervorden-Spatz syndrome has been designated Neurodegeneration with Brain Iron Accumulation, Type 1 (NBIA1).

CLINICAL CLASSIFICATION

The clinical diagnosis of HSS is based on criteria published by Swaiman (1991). A diagnosis of HSS was assigned only when a patient had all of the obligatory features, two or more of the corroborative features and none of the exclusionary features. The obligatory features include onset during the first two decades of life with progression of signs and symptoms. Extrapyramidal dysfunction is also present, including one or more of the following: dystonia, rigidity, choreoathetosis. Corroborative features include: corticospinal tract involvement, retinitis pigmentosa and/or optic atrophy, seizures, positive family history consistent with autosomal recessive inheritance, hypodense areas involving the basal ganglia on magnetic resonance imaging (MRI), and abnormal cytosomes in circulating lymphocytes and/or sea-blue histiocytes in bone marrow. Exclusionary features include abnormalities of ceruloplasmin or copper metabolism, overt neuronal ceroid-lipofuscinosis, severe retinal degeneration or visual impairment preceding other symptoms, family history of Huntington chorea or other dominantly inherited movement disorder, caudate atrophy, deficiency of β -hexosaminidase A or G_{M1}-galactosidase, nonprogressive course, and absence of extrapyramidal signs.

The symptoms of HSS typically begin with dystonia, followed by dysphagia and dysarthria (Hallervorden and Spatz, 1922; OMIM #124200). Patients often become rigid within 2-3 years of onset. Dystonia is a neurological movement disorder that produces turning or twisting movements, which may be repetitive or sustained, causing abnormal postures. Dystonia may affect one or many parts of the body. Like Hallervorden-Spatz syndrome, many neurodegenerative diseases, such as Parkinson's disease, Huntington's disease, and Wilson's disease, have features of dystonia (OMIM #556500, #143100, and

#277900, respectively). In HSS, as in these other disorders, dystonia is varied and seems to be a symptom, not a cause. These dystonic symptoms are generally ascribed to dysfunction in part of the brain called the basal ganglia. Idiopathic dystonia (or primary torsion dystonia; OMIM #128100) is characterized by age and site of onset, factors which play a role in the progression of the symptoms. The younger the age of onset, the more likely the dystonia will begin in one of the limbs, spread to other parts, and possibly become generalized. The older the age of onset, the more likely the dystonia will begin in the neck or cranial muscles and remain local. There is no alteration of consciousness, intellect, or sensation in idiopathic dystonia. Idiopathic dystonia is found in all ethnic and racial groups. It is distinguished by the lack of other neurological involvement, a strong hereditary predisposition, and the absence of specific neuropathology.

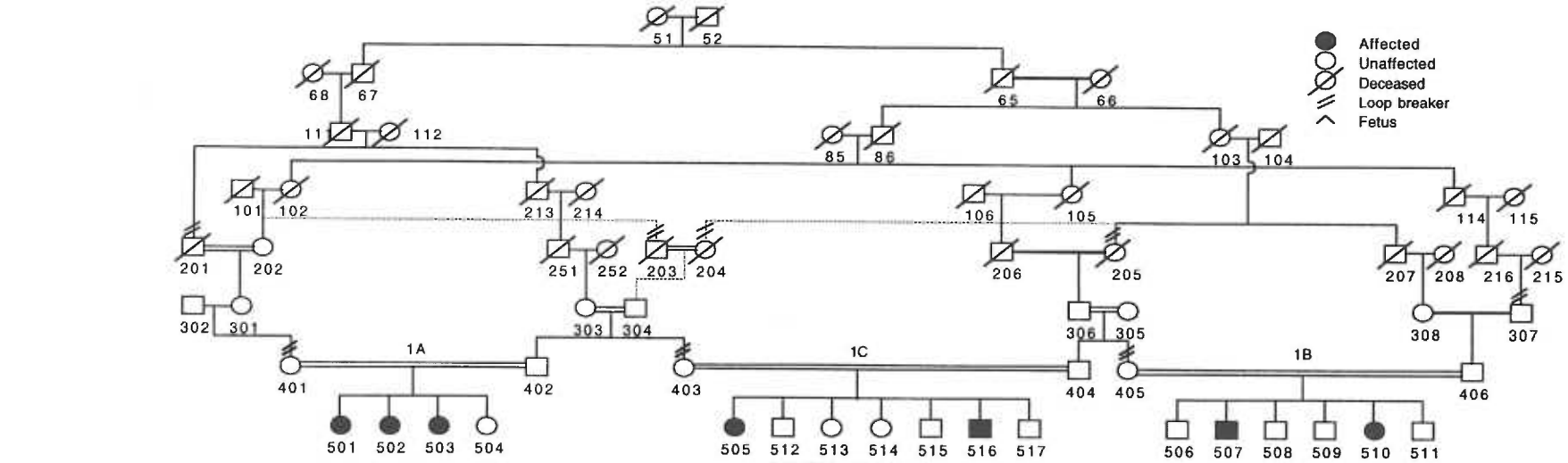
Families in this study were ascertained through clinical genetics or neurological centers. A neurologist or a medical geneticist evaluated all affected individuals. Detailed medical history, physical and laboratory findings, and pedigree data were taken for each family member. The diagnosis of HSS was assigned according to the outlined criteria. Information regarding the families participating in this study is summarized in **Table 1**. Pedigrees and haplotypes of all classical HSS families are shown in **Figure 1**.

Clinical classification of individuals was critical for this study. Misdiagnosis of disease phenotype can lead to over estimation of the recombination fraction between the disease and marker loci, and real linkage can be overlooked if the misclassification is extensive. For this reason, it was critical that all the families in our study fit the criteria of 'classical' Hallervorden-Spatz syndrome. Only the families that fit Swaimans' criteria and that demonstrated autosomal recessive inheritance were classified as classical HSS

phenotype. The average age of onset in our classical families was less than 10 years. Individuals with most of the features of HSS, but with late-onset (after the second decade) were classified as 'atypical' HSS phenotype. 'Probable' HSS phenotype meant that a patient likely had classical HSS but not enough information about the patient was available to make the distinction. A phenotype of 'unknown' meant that no information was sent with the patients' blood or DNA sample. The distribution of the phenotypes of the 59 families in this study is as follows: 48% classical HSS, 32% atypical HSS, 10% probable HSS, and 10% unknown HSS. Only the 28 classical HSS phenotype families were used in the linkage mapping part of this study. Of these 28 families, three were from Canada, four were from the USA, three were from Spain, seven families were from Italy, and 11 families were from several other countries. Families from the same countries were from varied locations. Though all of the Italian families were identified through one institute, their geographical distribution is unknown.

Family	Phenotype	Country	Linked to 20p13	# Affected	Consanguineous
HS1	Classical	USA	Yes	7	Yes
HS2	Classical	Canada	Insufficient data	1	No
HS5	Classical	New Zealand	Yes	1	No
HS7	Classical	Germany	Yes	1	No
HS8	Classical	Spain	Yes	1	No
HS9	Classical	Japan	No	1	Yes
HS10	Classical	Italy	Yes	2	Yes
HS11	Classical	Italy	Yes	2	No
HS12	Classical	Italy	Yes	1	No
HS13	Classical	Italy	Yes	3	No
HS14	Classical	Australia	Yes	1	No
HS15	Classical	Spain	Yes	1	No
HS16	Classical	Italy	Yes	2	No
HS19	Classical	Italy	Yes	1	No
HS20	Classical	Italy	Yes	2	No
HS23	Classical	USA	Insufficient data	1	Unknown
HS26	Classical	USA	Yes	1	No
HS27	Classical	Germany	Insufficient data	1	No
HS35	Classical	France	Yes	1	Yes
HS36	Classical	USA	Insufficient data	1	Unknown
HS37	Classical	Poland	Yes	2	No
HS48	Classical	Spain	Yes	1	Unknown
HS49	Classical	Saudi Arabia	No	3	Likely
HS50	Classical	Saudi Arabia	Insufficient data	1	Likely
HS54	Classical	Canada	Yes	1	No
HS55	Classical	Canada	No	2	Unknown
HS56	Classical	Canada	Insufficient data	1	Likely
HS60	Classical	Turkey	Yes	2	Yes
HS21	Probable	USA	Insufficient data	1	No
HS29	Probable	USA	Insufficient data	1	Unknown
HS43	Probable	Spain	Yes	1	Unknown
HS44	Probable	Spain	Yes	1	Unknown
HS45	Probable	Denmark	Yes	1	No
HS58	Probable	Holland	Yes	1	Unknown
HS3	Unknown	Unknown	Insufficient data	1	Unknown
HS34	Unknown	USA	Insufficient data	1	Unknown
HS41	Unknown	USA	No	1	Unknown
HS42	Unknown	USA	Yes	1	Unknown
HS57	Unknown	USA	Insufficient data	1	Unknown
HS59	Unknown	Middle East	Yes	2	Unknown
HS4	Atypical	USA	Yes	2	Yes
HS6	Atypical	USA	Insufficient data	1	Unknown
HS17	Atypical	USA	Insufficient data	1	No
HS18	Atypical	USA	Insufficient data	1	No
HS22	Atypical	USA	Yes	1	No
HS25	Atypical	USA	No	2	No
HS28	Atypical	USA	Insufficient data	1	Unknown
HS30	Atypical	Unknown	Insufficient data	1	Unknown
HS31	Atypical	UK	No	4	Yes
HS32	Atypical	USA	No	3	No
HS33	Atypical	USA	No	2	Unknown
HS38	Atypical	USA	Yes	1	No
HS39	Atypical	USA	Yes	2	No
HS40	Atypical	USA	Yes	2	Unknown
HS46	Atypical	Italy	No	3	No
HS47	Atypical	Germany	Yes	1	Unknown
HS51	Atypical	USA	Insufficient data	1	Unknown
HS52	Atypical	USA	Insufficient data	1	Unknown
HS53	Atypical	Scotland	Insufficient data	1	No

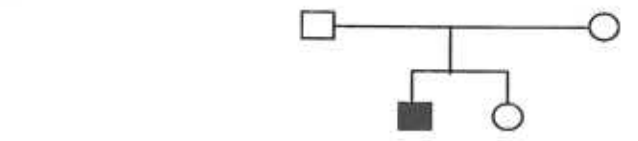
Table 1. HSS Family Summary.



HS1 USA - Amish

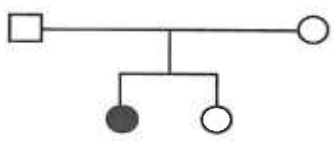
cM	Marker	302	301	401	501	502	503	504	402	303	304	403	505	512	513	514	515	516	517	404	405	506	507	508	509	510	511	406
	D20S906	6 6	1 4	1 6	1 1	1 1	1 1	6 1	1 4	1 4	4 4	1 4	1 3	4 3	1 4	1 4	4 4	1 3	4 4	3 4	3 4	4 3	3 1	4 1	4 1	3 1	4 1	3 1
	D20S179	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 4	2 2	2 2	2 2	2 2	2 6	2 6	2 6	2 2	2 6	2 6	2 6	6 2	2 2	6 2	6 2	2 2	6 2	2 2
	D20S113	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 5	1 5	1 5	1 1	1 5	1 5	1 5	5 1	1 1	5 1	5 1	1 1	5 1	1 1
1	GAAT4E12	2 2	1 2	1 2	1 1	1 1	1 1	2 1	1 2	1 2	2 2	1 2	1 1	2 1	1 1	1 1	2 1	1 1	2 1	1 1	1 1	1 2	1 1	1 1	1 1	1 1	1 1	2 1
	AFM074wa9	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2
	D20S198	2 3	3 6	3 2	3 3	3 3	3 3	2 3	3 6	3 6	6 6	3 6	3 3	6 3	3 7	3 7	6 7	3 3	6 7	3 7	3 7	7 3	3 3	7 3	7 3	3 3	7 3	3 3
1	D20S842	5 7	2 8	2 5	2 2	2 2	2 2	5 2	2 8	2 6	8 8	2 8	2 2	8 2	2 3	2 3	8 3	2 2	8 3	2 3	2 3	3 5	2 2	3 2	3 2	2 2	3 2	5 2
	AFMa049yd1	6 5	7 5	7 6	7 7	7 7	7 7	6 7	7 5	7 1	5 5	7 5	7 7	5 7	7 3	7 3	5 3	7 7	5 3	7 3	7 3	3 3	7 7	3 7	3 7	7 7	3 7	3 7
	D20S181	1 5	3 6	3 1	3 3	3 3	3 3	1 3	3 6	3 3	6 6	3 6	3 3	6 3	3 3	3 3	6 3	3 3	6 3	3 3	3 3	3 3	3 3	3 3	3 3	3 3	3 3	3 3
2	D20S193	5 3	3 4	3 5	3 3	3 3	3 3	5 4	3 4	3 5	4 4	3 4	3 3	4 3	3 3	3 3	4 3	3 3	4 3	3 3	3 3	3 1	3 3	3 3	3 3	3 3	3 3	1 3
	D20S473	4 3	4 3	4 4	4 4	4 4	4 4	4 3	4 3	4 6	3 3	4 3	4 4	3 4	4 5	4 5	3 5	4 4	3 5	4 5	4 5	5 1	4 4	5 4	5 4	4 4	5 4	1 4
	D20S867	5 3	1 1	1 5	1 1	1 1	1 1	5 1	1 1	1 2	1 1	1 1	1 1	1 1	1 4	1 4	1 4	1 1	1 4	1 4	1 4	4 4	1 1	4 1	4 1	1 1	4 1	4 1
	D20S116	3 1	1 4	1 3	1 1	1 1	1 1	3 4	1 4	1 4	4 4	1 4	1 1	4 1	1 1	1 1	4 1	1 1	4 1	1 1	1 1	1 5	1 1	1 1	1 1	1 1	1 1	5 1
	D20S889	7 9	5 5	5 7	5 5	5 5	5 5	7 5	5 5	5 3	5 5	5 5	5 5	5 5	5 11	5 11	5 11	5 5	5 11	5 11	5 11	11 11	5 5	11 5	11 5	5 5	11 5	11 5

Figure 1. HSS family pedigrees are shown with the disease-associated haplotype (shaded). In order to run the LINKAGE/FASTLINK programs for family HS1, inbreeding loops were broken as indicated. Dashed lines indicate relationships that were not considered because they do not significantly affect linkage calculations. Minus signs indicate genotypes that have not been determined.



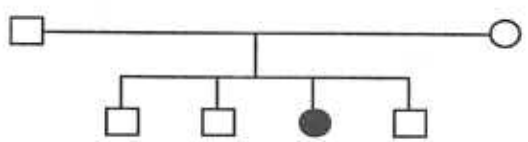
HS5 New Zealand

cM	Marker	201		301		302		202	
	D20S906	4	3	4	5	3	6	6	5
	D20S179	4	2	4	2	2	2	2	2
	D20S113	1	1	1	5	1	1	1	5
1	GAAT4E12	2	2	2	2	2	2	2	2
	AFM074wa9	2	1	2	2	1	1	1	2
	D20S198	1	6	1	3	6	2	2	3
1	D20S842	3	8	3	6	8	5	5	6
	AFMa049yd1	5	3	5	3	3	3	3	3
	D20S181	5	5	5	3	5	5	5	3
2	D20S193	5	4	5	3	4	5	5	3
	D20S473	4	4	4	5	4	4	4	5
	D20S867	6	1	6	5	1	3	3	5
	D20S116	3	4	3	2	4	6	6	2
	D20S889	7	5	7	1	5	5	5	1



HS7 Germany

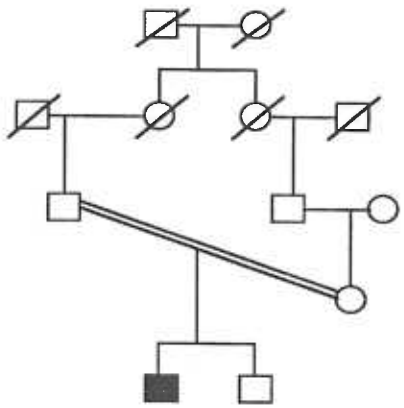
101		201		202		102	
3	3	3	4	3	6	6	4
-	-	2	2	-	-	-	-
5	1	1	1	1	3	3	1
1	2	2	1	2	1	1	1
1	2	2	2	2	2	2	2
7	1	1	1	1	4	4	1
1	3	3	3	3	3	3	3
5	7	7	6	7	6	6	6
7	3	3	5	3	5	5	5
6	3	3	3	3	5	5	3
2	2	2	4	2	4	4	4
1	3	3	2	3	4	4	2
1	1	1	1	1	4	4	1
5	12	12	2	12	2	5	2

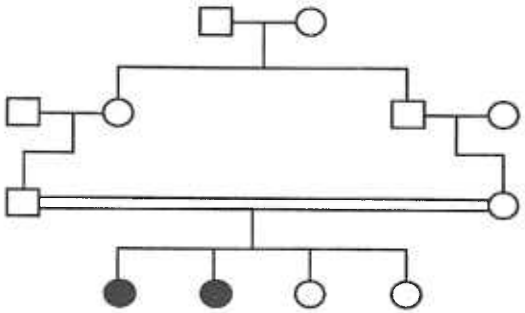


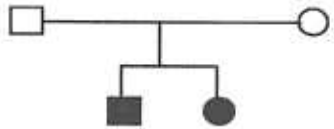
HS8 Spain

201		301		302		303		304		202	
1	3	1	2	3	5	1	5	1	2	2	5
2	2	2	2	2	2	2	2	2	2	2	2
2	1	2	1	1	1	2	1	2	1	1	1
2	1	2	2	1	2	2	2	2	2	2	2
2	1	2	1	1	2	2	2	2	1	1	2
3	2	3	2	2	7	3	7	3	2	2	7
4	11	4	4	11	8	4	8	4	4	4	8
2	2	2	8	2	3	2	3	2	8	8	3
5	6	5	5	6	5	5	5	5	5	5	5
5	3	5	5	3	5	5	5	5	5	5	5
5	3	5	2	3	5	5	5	5	2	2	5
2	3	2	4	3	3	2	3	2	4	4	3
4	3	4	1	3	3	4	3	4	1	1	3
6	5	6	1	5	13	6	13	6	1	1	13

Figure 1. (continued)







cM	Marker	HS9 Japan 301		501		502		401	
	D20S906	4	3	4	3	4	3	3	3
	D20S179	2	1	2	5	2	5	2	5
	D20S113	3	3	3	6	3	6	1	6
1	GAAT4E12	1	2	1	1	1	1	1	1
	AFM074wa9	3	2	3	2	3	2	2	2
	D20S198	1	3	1	3	1	3	2	3
1	D20S842	3	4	3	8	3	8	4	8
	AFMa049yd1	2	2	2	3	2	3	2	3
	D20S181	3	3	3	3	3	3	8	3
2	D20S193	3	7	3	2	3	2	6	2
	D20S473	4	5	4	4	4	4	4	4
	D20S867	4	4	4	6	4	6	1	6
	D20S116	5	6	5	6	5	6	4	6
	D20S889	5	5	5	1	5	1	8	1

HS10 Italy (101)		201		202		203		204		(102)	
3	4	3	3	3	3	4	6	3	3	6	3
7	2	7	7	7	7	2	2	2	7	2	7
5	1	5	5	5	5	1	1	1	5	1	5
1	2	1	1	1	1	2	2	2	1	2	1
2	1	2	2	2	2	1	3	1	2	3	2
3	1	3	3	3	3	1	3	1	3	3	3
3	1	3	3	3	3	1	4	1	3	4	3
5	2	5	5	5	5	2	5	2	5	5	5
5	2	5	5	5	5	2	1	2	5	1	5
5	4	5	5	5	5	4	4	4	5	4	5
3	3	3	3	3	3	3	4	3	3	4	3
1	1	1	1	1	1	1	3	1	1	3	1
3	3	3	3	3	3	3	3	3	3	3	3
7	5	7	7	7	7	5	5	5	7	5	7

HS11 Italy 101		201		202		102	
3	3	3	3	3	3	3	3
2	3	2	3	2	3	2	3
1	3	1	1	1	1	6	1
2	1	2	2	2	2	2	2
1	2	1	2	1	2	2	2
6	1	6	3	6	3	6	3
7	9	7	5	7	5	13	5
3	6	3	5	3	5	4	5
3	1	3	1	3	1	3	1
6	3	6	3	6	3	5	3
4	1	4	1	4	1	4	1
1	3	1	2	1	2	2	2
1	4	1	2	1	2	3	2
2	5	2	5	2	5	5	5

Figure 1. (continued)

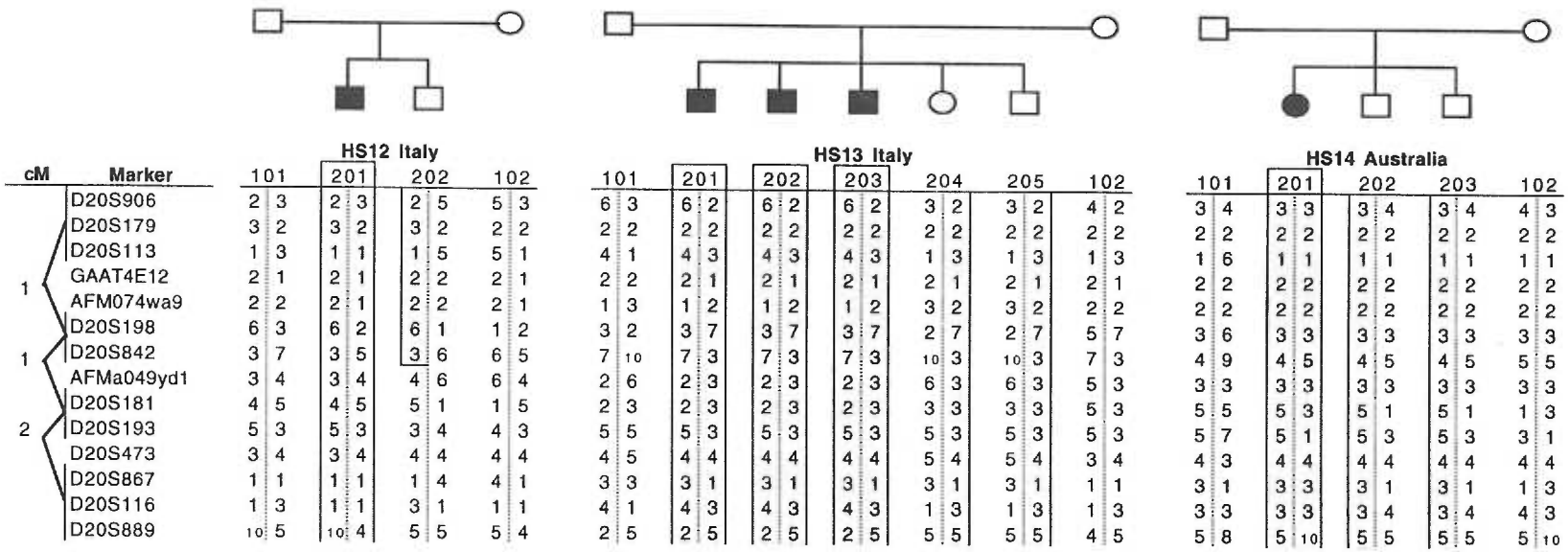


Figure 1. (continued)

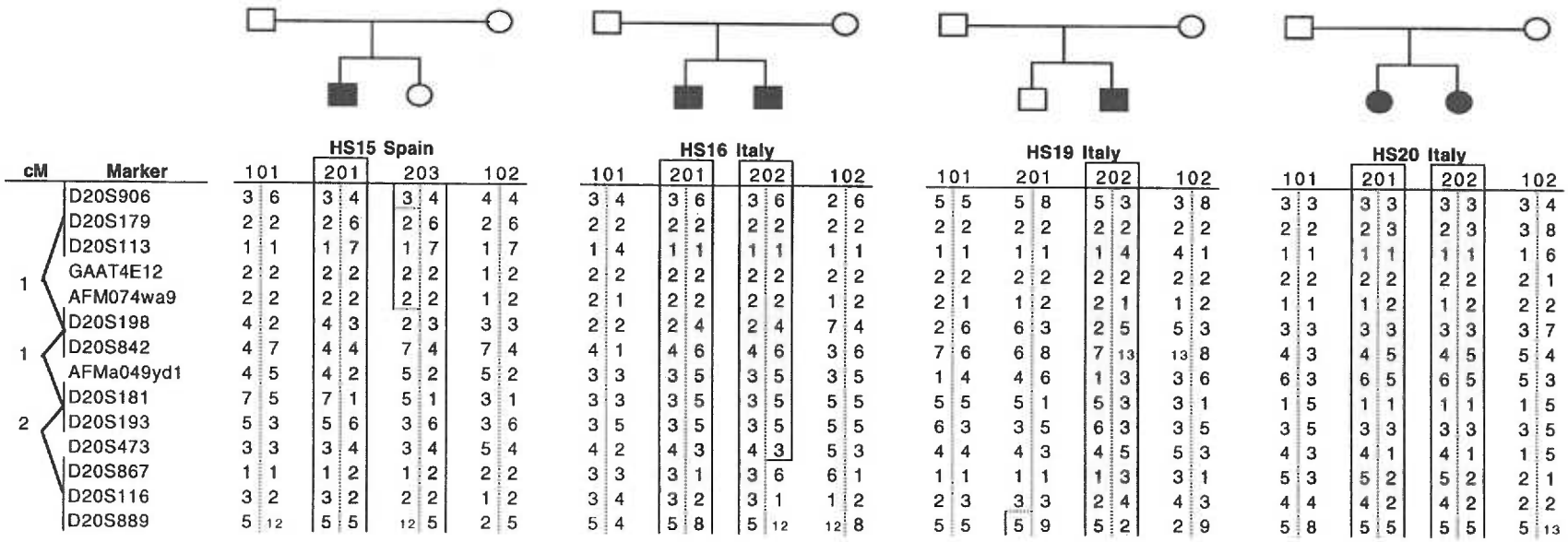
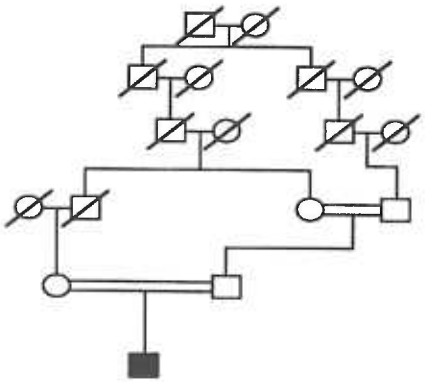
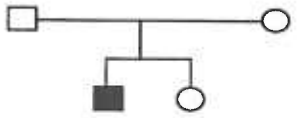
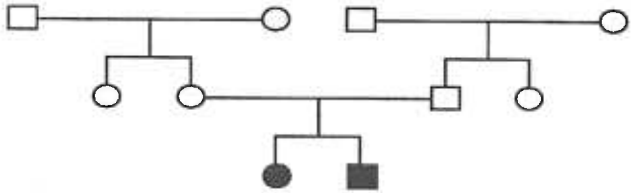


Figure 1. (continued)







cM	Marker	HS26 USA 101		201		202		102		HS35 Algeria 501		601		502		403		404		HS37 Poland 102		201		202		301		302		203		204		103	
	D20S906	1	4	1	3	1	6	3	6	6	2	2	6	6	7	6	3	7	5	4	4	4	4	4	4	4	5	4	5	5	3	4	4	4	3
	D20S179	2	2	2	2	2	2	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	D20S113	3	1	3	1	3	1	1	1	1	1	1	1	1	6	1	7	6	5	3	1	1	1	3	1	1	1	1	1	1	1	5	1	1	1
1	GAAT4E12	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	3	2	1	2	1	1	1	2	1	1	2	1	2	2	2	2	2	2	2
	AFM074wa9	2	2	2	1	2	2	1	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2
	D20S198	2	5	2	6	2	3	6	3	3	3	3	3	3	3	3	3	3	3	2	2	2	6	2	3	3	6	3	6	6	3	3	7	7	3
1	D20S842	5	2	5	4	5	4	4	4	14	12	12	14	14	7	14	7	7	6	7	5	5	3	5	4	4	3	4	3	3	7	-	-	5	7
	AFMa049yd1	5	4	5	5	5	3	5	3	2	6	6	2	2	3	2	2	3	6	5	5	5	3	5	5	5	2	5	2	2	5	3	3	3	5
	D20S181	5	1	5	1	5	3	1	3	6	4	4	4	4	5	4	2	5	3	5	5	5	5	5	3	3	3	3	3	3	3	5	3	3	3
2	D20S193	5	3	5	3	5	3	3	3	3	3	3	3	3	14	3	5	14	3	5	4	4	5	5	3	3	3	3	3	3	3	5	3	3	3
	D20S473	4	1	4	1	4	4	1	4	3	4	4	4	4	2	4	4	2	4	3	4	4	3	3	5	5	1	5	1	1	1	3	4	4	1
	D20S867	1	2	1	2	1	3	2	3	3	1	1	1	1	6	1	3	6	4	1	1	1	6	1	2	2	1	2	1	1	3	1	4	4	3
	D20S116	2	4	2	2	2	3	2	3	3	3	3	3	3	5	3	3	5	8	1	1	1	3	1	3	3	3	3	3	3	5	5	9	9	5
	D20S889	14	2	14	5	14	5	5	5	9	6	6	6	6	11	6	12	11	13	8	8	8	10	8	11	11	13	11	13	-	-	5	16	-	-

Figure 1. (continued)

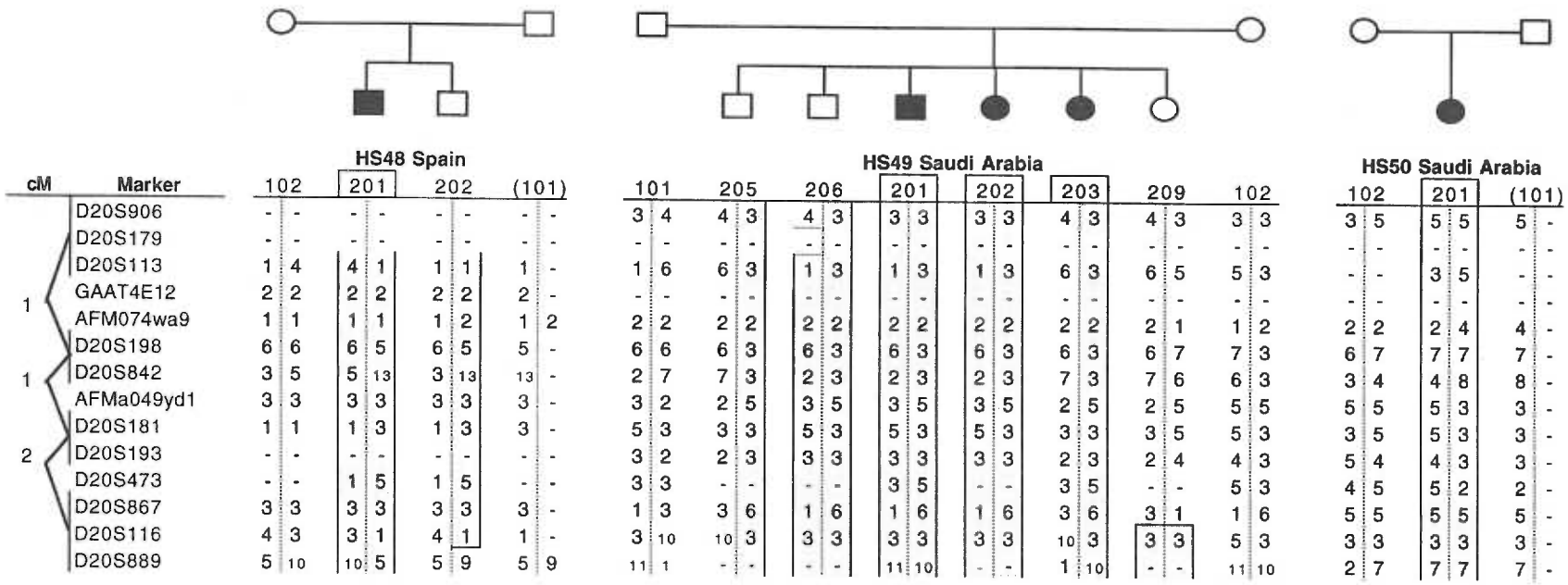


Figure 1. (continued)

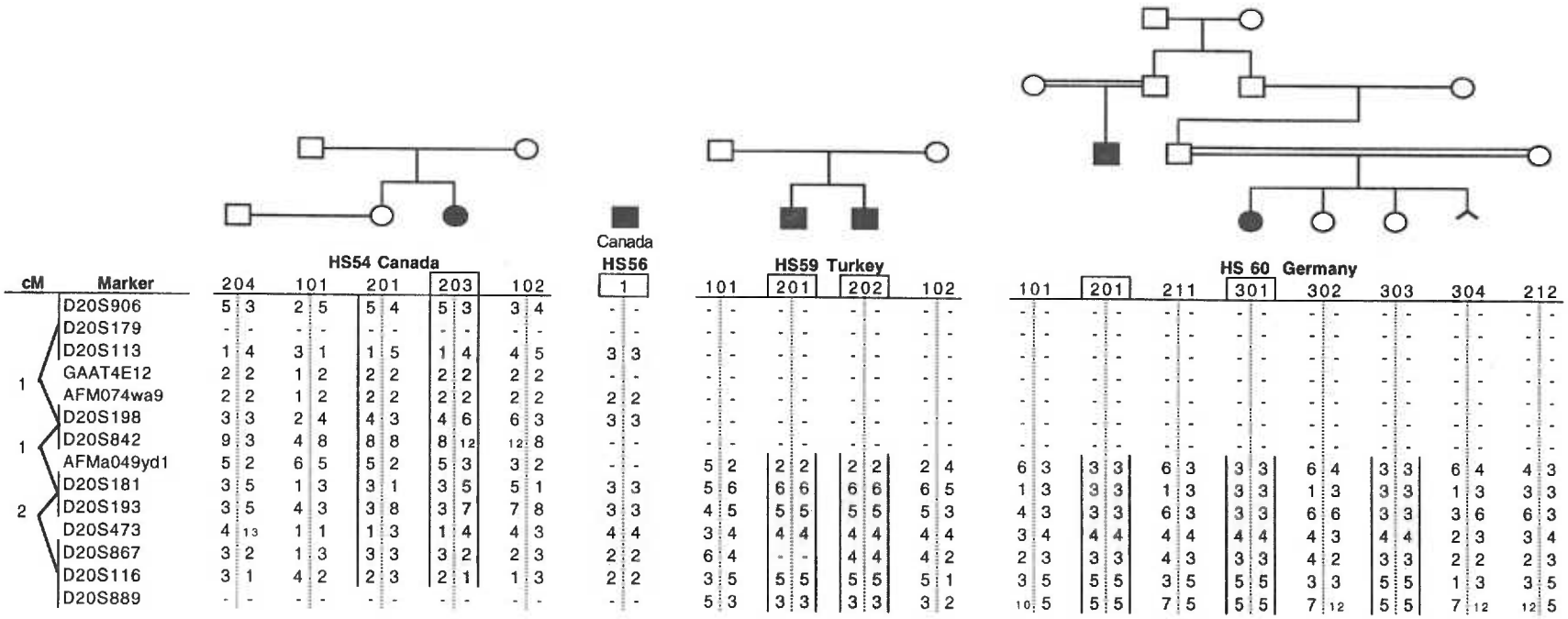


Figure 1. (continued)

PATHOGENESIS

HSS is characterized pathologically by iron accumulation and axonal spheroid formation in the basal ganglia, particularly in the globus pallidus and pars reticulata of the substantia nigra. The fundamental defect responsible for (or leading to) the brain iron accumulation remains to be elucidated.

Julius Hallervorden and Hugo Spatz first described the clinical and pathologic features of this disorder in 1922 (Hallervorden and Spatz, 1922). They reported five sisters with gait abnormalities, dysarthria, dementia, and the striking gross pathologic finding of rust-brown pigmentation of the globus pallidus (GP) and the reticular zone of the substantia nigra (SNR). Iron is the major component of this pigment (Hallervorden, 1924). Ceroid-lipofuscin and neuromelanin also accumulate both intra- and extraneuronally in the brains of people with HSS. Other neuropathologic findings include demyelination, neuronal loss, and gliosis. This damage occurs predominantly within the GP and SNR, where focal, symmetric destruction is grossly evident. Magnetic resonance imaging (MRI) of brain in patients with HSS shows pallidal iron accumulation (Drayer, 1987; Tanfani et al., 1987; Mutoh et al., 1988; Sethi et al., 1988). Serial MRI studies indicate an accumulation of iron over time (Gallucci et al., 1989).

Additional, cytologic abnormalities, not specific to this disorder, are occasionally reported in patients with HSS. These findings include sea-blue histiocytes, bone marrow macrophages (which contain clumped storage material), circulation lymphocytes (with vacuoles and cytosomic bodies; Swaiman et al., 1983), and acanthocytes (Roth et al., 1971; Swisher et al., 1972; Luckenbach et al., 1983). The storage material in histiocytes

is probably ceroid-lipofuscin, and the lymphocytic cytosomes are similar to those seen in ceroid-lipofuscin storage disorders. Acanthocytosis has been reported in a subset of patients with HSS (Tripathi et al., 1992; Higgins et al., 1992). Lipofuscin accumulation and acanthocyte formation are secondary to lipid peroxidation, a process stimulated by iron (Defendini et al., 1973; Park et al., 1975).

In normal human brain iron distribution is highly regionalized, with iron accumulation being greatest in globus pallidus and substantia nigra pars reticulata (Spatz, 1922; Spatz, 1922). The pattern of excessive iron accumulation in HSS parallels the pattern of normal iron accumulation in unaffected individuals. In HSS, the globus pallidus and substantia nigra contain two to three times the normal amount of iron (Kornyey, 1964; Volkl and Ule, 1972; Vakili et al., 1977), yet iron content is normal in other regions of the brain (Kornyey, 1964). Peripheral iron homeostasis is normal, as are CSF levels of iron, transferrin and ferritin (Szanto and Gallyas, 1966; Swaiman et al., 1983). In patients with HSS, *in vivo* ^{59}Fe-uptake studies indicate that accumulation of iron in the basal ganglia is secondary to increased iron uptake with normal turnover (Szanto and Gallyas, 1966; Vakili et al., 1977; Swaiman et al., 1983).

Normal brain iron accumulation is related to the metabolism of gamma-aminobutyric acid (GABA), the major inhibitory neurotransmitter in the vertebral central nervous system (CNS). Glutamate is a precursor for GABA metabolism, and glutamate-binding proteins contain iron (Michaelis et al., 1982). Iron is also required for succinate dehydrogenase, the enzyme through which the product of the GABA shunt re-enters the tricarboxylic acid cycle. In normal mammalian brain, high iron content in the basal ganglia coincides with GABAergic neuronal pathways (Hill and Switzer, 1984), and iron

is released from these neurons under normal physiologic conditions. Inhibition of GABA catabolism in the GP results in reduced histochemical iron in the ipsilateral substantia nigra (Hill, 1985). Conversely, unilateral microinjection of excitatory amino acids into the ventral striatum leads to increase in histochemical iron concentration in the ipsilateral ventral pallidum, globus pallidus, and substantia nigra pars reticulata (Shoham et al., 1992).

The balance between excitatory and inhibitory amino acids may be critical in maintaining normal brain iron. Perry and colleagues (1985) demonstrated that GABA content was significantly decreased in the globus pallidus and substantia nigra. Gamma-vinyl GABA (GVG) is an enzyme-activated irreversible inhibitor of GABA transaminase (the enzyme responsible for the degradation of GABA) which when injected into iron-rich regions of rat brain results in a significant reduction in iron concentration (Hill, 1985). In mammalian brain, glutamic acid decarboxylase (GAD) catalyzes the conversion of glutamic acid to GABA. These findings suggest iron accumulation in the basal ganglia may be related to altered GABA metabolism or transport.

Two major active proteins involved in iron homeostasis are transferrin and ferritin. Transferrin is an iron transport protein that binds to a specific plasma membrane receptor. It enters the cell and deposits iron in the endosomes. In brain, the transferrin receptor gene is negatively regulated by iron and may have a function in myelination (Connor, 1986). Transferrin is ubiquitously distributed on the surface of actively growing human cells, but is concentrated in oligodendrocytes in normal brain (Hill, 1984; Connor, 1986; Connor, 1987). Ferritin is the major intracellular iron storage protein in all organisms. Its binding in brain, as in other tissues, is the most efficient mechanism of iron storage that

permits mobilization of intracellular iron on demand. Ferritin converts ferrous iron to its ferric form and retains it in an inactive state until it is needed in intracellular metabolism, thus circumventing the peroxidation potential of iron.

Iron initiates a variety of free radical reactions in mammals and is able to induce damage to membranes by peroxidation (Halliwell and Gutteridge, 1984 and 1985), the chain reaction in which polyunsaturated fatty acids undergo oxidative destruction. Iron accumulation is well documented in HSS, and many of the pathologic features of this disease can be explained based on lipid peroxidation. Lipofuscin and melanolipofuscin, degradation products of lipid peroxidation (Defendini et al., 1973; Park et al., 1975), are abundant in a variety of cell types in HSS. Ferrous iron leads to production of lipid peroxides, disappearance of long chain polyunsaturated fatty acids, and progressive photoreceptor degeneration (Rapp et al., 1977). There is much experimental and pathological evidence that iron initiates lipid peroxidation, which, in turn, causes many of the pathologic features of HSS. The question remains whether iron accumulation in HSS is primary, or whether it represents an intermediate step in a complex, neuropathologic cascade in which the primary defect remains to be identified.

Identification of the gene for HSS may clarify the mechanism by which iron accumulates normally and in some common neurodegenerative disorders. Increased brain iron is recognized in Parkinson disease (Sofic et al., 1988; Dexter et al., 1989; Riederer et al., 1989; Sofic et al., 1991; Hirsch et al., 1991), Alzheimer disease (Goodman, 1953), and Huntington disease (Jackson et al., 1972) and tardive dyskinesia (Ben-Schachar et al., 1994). Iron accumulation in these disorders is likely a secondary phenomenon, with neuronal damage following, in part, from iron-catalyzed oxidation. Iron functions in the

formation and maintenance of myelin sheet, and may play a role in the pathology of myelin diseases (LeVine and Macklin, 1990).

Retinitis pigmentosa (RP) is not an obligatory feature of Hallervorden-Spatz syndrome; however, Newell et al. (1979) suggested that HSS may comprise two distinct disorders, one of which is combined with pigmentary degeneration of the retina. Their review of the more than 40 patients with HSS reported by Dooling et al. (1974) revealed that individuals with retinitis pigmentosa have an earlier onset and more rapid course with earlier death compared to individuals without this feature. GABA is the predominant neurotransmitter in mammalian retina, yet no association is recognized between iron metabolism and GABA in retina, and no iron accumulates in HSS retina. Studies in a feline model of RP showed gradual selective loss of GABA immunoreactivity in both the inner and outer parts of the retina (Ehinger et al., 1991).

There is currently no mouse or rat model for Hallervorden-Spatz syndrome. The mouse and rat genome databases were searched for all of the phenotypic traits associated with HSS. No strains were detected with iron accumulation in the basal ganglia or any other of the striking features of HSS. The only strains found with any interesting similarities to HSS were the WAG rat (Origin: AL Bacharach, Glaxo Ltd. 1924 from Wistar stock) and mice homozygous for the rod photoreceptor cGMP-phosphodiesterase *rd* mutation (OMIM #180072). The WAG rat has low brain GABA (Rick et al., 1971) and bilateral retinal degeneration with early onset and a slowly progressive course, which resembles human retinitis pigmentosa (Lai et al 1980). Mice homozygous for the *rd* mutation display hereditary retinal degeneration, which has been considered a model for human retinitis pigmentosa. In affected animals, the retinal rod photoreceptor cells begin

degenerating at about postnatal day 8, and by 4 weeks no photoreceptors are left. Farber and Lolley (1974, 1976) showed that this degeneration is preceded by accumulation of cyclic GMP in the retina and is correlated with deficient activity of the rod photoreceptor cGMP-phosphodiesterase.

Despite detailed pathologic and physiologic investigations, the pathogenesis of Hallervorden-Spatz syndrome remains unclear. Tenable theories of pathogenesis have been proposed (Park et al., 1975; Perry et al., 1985; Swaiman, 1991; Tripathi et al., 1992) based on obstructive sequelae of iron accumulation and lipid peroxidation. Though it is not established that either of these processes is primary, they form a basis for identifying candidate genes for this disorder. Genes encoding proteins for GABA and iron metabolism, and the proteins they interact with, as well as loci associated with isolated RP and other forms of syndromic RP can be considered candidate genes for HSS.

MATERIALS AND METHODS

DNA preparation. Informed consent was obtained from all participants. Fifteen milliliters of whole blood was obtained, and DNA was isolated by standard procedures (Miller et al., 1988). Additional samples were drawn for some individuals for Epstein-Barr virus lymphoblastoid transformation. Lymphocytes were either immediately transformed or frozen for subsequent transformation (Louie and King, 1991). Cultured fibroblasts were obtained from a deceased individual with HSS. Paraffin-embedded brain, including globus pallidus and substantia nigra pars reticulata, was obtained from a man with atypical HSS.

One affected member of the HS1 family has been karyotyped at the 650-band level with normal results by the Cytogenetics Laboratory at the Children's Hospital of Buffalo. Chromosome rearrangements and deletions, while useful in the localization of many autosomal dominant and X-linked genes, are exceedingly rare in autosomal recessive disorders and have not been reported in HSS.

Marker selection. A genetic marker can be defined as an observable polymorphism within the study population. Since Botstein and colleagues (1980) first described restriction fragment length polymorphisms as a new class of genetic marker, many other types of markers have been identified, namely VNTRs (variable number of tandem repeats) and SSLPs (simple sequence length polymorphisms, also known as simple sequence repeats, SSRs). The most useful class of polymorphisms for genomic screening and fine genetic mapping is the SSLPs. Unlike standard RFLPs and VNTRs, SSLPs require only minute amounts of DNA, are generally easy to genotype by polymerase

chain reaction (PCR, Mullis and Faloona, 1987), are spread ubiquitously throughout the genome, and are highly polymorphic. The first and most widely used simple sequence repeats to be developed for genotyping by PCR assay were the simple $(CA)_n$ repeats (Litt and Luty, 1989; Weber and May, 1989; Weber, 1990). On average, variable $(CA)_n$ repeats occur every 40 to 50 kb (Weber and May, 1989) and highly variable repeats occur every 300 to 500 kb. Thousands of SSLP markers have now been described by Généthon (Weissenbach et al., 1992; Gyapay et al., 1994; Dib et al., 1996); the Cooperative Human Linkage Center (CHLC; Murray et al., 1994); and the Utah Marker Development Group (1995). Genotyping using SSLPs requires the careful selection of standards and controls. Standard DNA should be included in every gel to so that consistent allele sizes are given for all genotyped samples (usually CEPH individuals 1347-02, 1331-01, and 1331-02 are used as sources for standards; Dausset, 1990). Consistency is essential in linkage and linkage disequilibrium studies, since they are usually dependent on allele frequency estimates.

For the linkage analysis portion of this study, we used tandem repeat, microsatellite markers (Litt and Luty, 1989; Weber and May, 1989; Weber, 1990; Edwards et al., 1991). Most of the markers were obtained from Research Genetics, Inc. (Huntsville AL, USA). Initially, highly polymorphic markers were selected in autosomal regions known to contain candidate genes related to iron transport and metabolism, and loci associated with isolated or syndromic retinitis pigmentosa (**Table 2**). Dystonia is not present in all patients with HSS so loci associated with idiopathic dystonia were not considered as candidate genes for HSS.

Name	Abbreviation	Location	OMIM #
Ferritin heavy chain 1	FTH1	11q12-q13	134770
Ferritin light chain	FTL	19q13.3-q13.4	134790
Gamma aminobutyric acid transporter	GABATR	3p25-24	137165
Gamma aminobutyric acid-alpha 1	GABRA1	5q34-q35	137160
Gamma aminobutyric acid-alpha 2	GABRA2	4p13-p12	137140
Gamma aminobutyric acid-alpha 6	GABRA6	5q31.1-q35	137143
Gamma aminobutyric acid-beta 1	GABRB1	4p13-p12	137190
Gamma aminobutyric acid-beta 3	GABRB3	15q11.2-q12	137192
Gamma aminobutyric acid-gamma 2	GABRG2	5q31.1-q33.1	137164
Glutamic acid decarboxylase 1	GAD1	2q31	266100
Glutamic acid decarboxylase 2	GAD2	10p11.23	138275
Glutamic acid decarboxylase 3	GAD3	22q13	138276
Beta subunit of cGMP phosphodiesterase	PDEB	4p16.3	180072
Peripherin	PRPH	12q12-q13	170710
Rhodopsin	RHO	3q21-q24	180380
Retinitis pigmentosa 9	RP9	7p15.1-p13	180104
Retinitis pigmentosa 10	RP10	7q31-q35	180105
Transferrin	TF	3q21	190000
Transferrin receptor (p90)	TFRC	3q29	190010

Table 2. Candidate gene regions first screened for linkage to HSS.

PCR analysis. The microsatellite markers were used to amplify genomic DNA from the HSS families. PCR for most markers was done in a total volume of 7 µl, with 14 ng genomic DNA, 3.5 pmol of each primer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 50 mM KCl, 10 mM Tris-Cl, pH 8.5, and 0.175 U of Ampli*Taq* DNA Polymerase (Perkin-Elmer Cetus Corp., Norwalk CT, USA). In a typical experiment, three-temperature "touchdown" PCR amplification (Don et al., 1991) was performed in a Gene Amp PCR System 9600 thermocycler (Perkin-Elmer Cetus Corp.), using an initial annealing temperature of 65°C which was decreased by one degree in each of the first fifteen cycles, then maintained at 50°C for 20 more cycles. Each cycle consisted of a 15 second 94°C denaturing step, a 30-second annealing step and a 30-second 72°C extension step. This was followed by a final 5-minute extension step at 72°C, after which the temperature was reduced to 15°C until the reaction tubes were removed from the thermocycler. An initial hold step at 94°C for 4-5 minutes was often used in order to maximize denaturation of genomic DNA.

Genotyping. After PCR amplification, an equal volume of 95% formamide loading buffer was added to each reaction mixture. The DNA products were resolved on denaturing 8% polyacrylamide gels containing 32% formamide and 5.6 M urea (Litt et al., 1993). The gels were pre-run for 30-45 minutes at 80W constant power, bringing the surface temperature of the front glass plate to approximately 50°C. Prior to loading, the samples were denatured at 95°C for 5 min. and allowed to cool to room temperature. Shark-tooth combs placed at the top of the gel were used for wells, allowing for up to 65 samples to be run per load. About 3 µl of each sample was loaded onto the gel and run for a predetermined length of time to insure proper separation. The sample run-time was based on the rate at which the DNA fragments migrate versus the rate at which the xylene

in the buffer migrates. When planning a gel run, the range of fragment sizes for each marker, the space between loads (1-2 cm), and the number of loads on the gel must be taken into consideration. Several markers of known fragment size were run and the distances they migrated were measured. This information was utilized to construct a chart for estimating the rates of migration for DNA fragments ranging from 70 to 500 bp (Protocols). The following equation was extrapolated from the data:

$$y = 5100x^{-0.8807}$$

where y equals the migration rate (versus xylene; percent) and x equals the marker fragment size (nt). Once the gel has finished running, the DNA fragments were transferred by capillary blotting overnight onto positively charged nylon membrane (Hybond-N+ version 2.0, Amersham Life Science, Uppsala, Sweden). Large membranes were cut into smaller, more manageable sizes and labeled for reassembly. The blots were fixed and hybridized to $(CA)_{15}$ end-labeled with $\gamma\ ^{32}$P-dATP (Current Protocols in Human Genetics, 1994) using polynucleotide kinase (New England BioLabs, Beverly MA, USA). The blots were then placed in saran wrap (to prevent them from drying out) and were exposed to X-ray film (X-OMAT AR, Kodak; Hyperfilm MP, Amersham Life Science). Time of exposure varied from 2 hours to several days depending on the intensity of the signal produced from the hybridized probe. A screen can be used to intensify the amount of exposure, but is not recommended unless time is a concern. When an intensifying screen was used, the bands were fuzzy and sometimes difficult to read, especially when space between alleles was minimal. An example of an autoradiograph is shown in **Figure 2**.

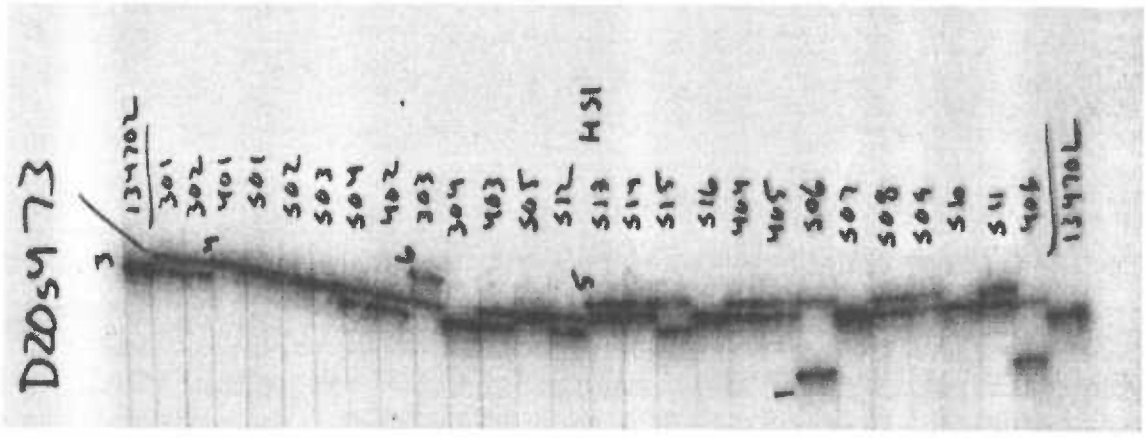


Figure 2. An autoradiograph of PCR products generated from the HS1 family using trinucleotide repeat marker D20S473. All the affected individuals (501, 502, 503, 505, 516, 507 and 510) are homozygous for allele 4. Sample 1347-02, on both ends, is a CEPH (Centre d'Etude du Polymorphisme Humain) reference individual (the size of allele 3 is 181 base pairs).

After the autoradiograph had been developed, the film was labeled with the current date, type of probe, exposure time, and marker names. The samples were labeled (individuals were given aliases that do not disclose their relation to one another) and the CEPH standard individuals genotypes were marked. Genotypes for each lane were read in accordance with the CEPH standards and the information was transferred to a computer database. We first used the relational database program Paradox (Borland, Scotts Valley CA, USA) to store genotype data, but eventually moved the information into an Access database (Microsoft, Redmond WA, USA). Storing information in a relational database made it easy to export the data for linkage and other analyses. When any errors were detected in the genotypic data, such as non-Mendelian inheritance, suspect data was re-read, re-genotyped, and when necessary, the individuals were re-sampled or eliminated from the analysis.

Although primer multiplexing using multiple primer pairs simultaneously can be an efficient method for genotyping large numbers of samples very rapidly using minimal amounts of DNA (Church and Kieffer-Higgins, 1988; Weissenbach et al., 1992; Vignal et al., 1993), we chose instead to run the primers separately. Markers with larger fragments were loaded onto the gel first allowing us to run up to 9 loads (with 9 different markers) per 37-cm polyacrylamide gel with uniform separation of alleles across the gel. The only drawbacks with this method were increased gel running time (8-10 hours) and frequent loading of the gel. Ease in scoring the markers and in planning the gels compensated for these drawbacks.

Mapping Strategy. We first took a direct screening approach. Based on the known biochemical, physiological, and pathological features of HSS, we identified several

candidate genes. These included genes encoding proteins for GABA and iron metabolism, the proteins they interact with, and genes associated with several forms of retinitis pigmentosa (**Table 2**). Highly polymorphic markers were selected near these candidate gene regions. Once these regions were excluded by linkage and haplotype analysis (**Table 3**), we undertook a random screen of the entire autosomal genome.

cM	Marker	Lod score at θ = 0.00	0.01	0.05	0.10	0.20	0.30	0.40	Candidate Genes
14	D2S141	−∞	-4.09	-2.17	-1.36	-0.65	-0.31	-0.12	GAD1
16	D2S294	−∞	-6.28	-2.41	-1.02	-0.05	0.2	0.17	
	D2S118	−∞	-5.87	-3.94	-2.63	-1.34	-0.67	-0.27	
9	D3S1307	−∞	-5.25	-2.65	-1.56	-0.61	-0.19	-0.01	GABATR
6	D3S1620	−∞	-5.35	-1.88	-0.44	0.37	0.37	0.18	
8	D3S1304	−∞	-1.51	-0.26	0.14	0.36	0.34	0.22	
6	D3S1597	−∞	-7.83	-4.27	-2.57	-1.19	-0.64	-0.32	
6	D3S1263	−∞	-3.60	-1.08	-0.21	0.31	0.35	0.22	
6	D3S1585	-4.87	-2.86	-2.4	-2.38	-1.89	-1.1	-0.5	
7	D3S1293	−∞	-5.04	-1.84	-0.69	0.09	0.26	0.20	
14	D3S1567	-8.48	-2.04	-0.8	-0.37	-0.07	0.01	0.01	
	D3S1298	−∞	-7.36	-3.31	-1.73	-0.46	0.00	0.11	
15	D3S1558	−∞	-3.6	-1.1	-0.3	0.2	0.3	0.1	-RHO
11	D3S1292	−∞	-6.23	-2.94	-1.70	-0.73	-0.34	-0.12	-TF
	D3S1569	−∞	-7.25	-3.47	-1.66	-0.32	0.03	0.02	
11	D3S1314	−∞	-3.3	-1.2	-0.2	0.3	0.3	0.2	-TFRC
	D3S1311	−∞	-7.95	-5.42	-3.85	-2.03	-1.10	-0.49	
12	D4S412	−∞	-3.67	-1.69	-0.93	-0.31	-0.06	0.03	PDEB
7	D4S394	−∞	-3.54	-1.57	-0.82	-0.24	-0.02	0.04	
	D4S1582	−∞	-6.78	-2.92	-1.52	-0.46	-0.08	0.04	
11	D4S418	−∞	-8.19	-4.30	-2.74	-1.36	-0.67	-0.25	GABRA2
9	D4S405	−∞	-9.85	-7.08	-4.81	-2.32	-1.12	-0.44	-GABRB1
	D4S428	−∞	-4.23	-2.32	-1.48	-0.72	-0.35	-0.13	
9	D5S436	−∞	-10.73	-6.21	-3.98	-1.74	-0.74	-0.23	GABRA6
8	D5S410	−∞	-2.97	-1.63	-1.09	-0.58	-0.31	-0.12	-GABRA1, GABRG2
12	D5S422	−∞	-6.18	-4.11	-2.67	-1.16	-0.46	-0.11	
	D5S400	−∞	-4.18	-3.43	-2.58	-1.37	-0.70	-0.28	
17	D7S493	−∞	-5.25	-2.67	-1.60	-0.67	-0.25	-0.06	-RP9
	D7S526	-9.64	-3.68	-1.72	-0.98	-0.38	-0.15	-0.05	
8	D7S523	−∞	-6.61	-3.69	-2.21	-0.97	-0.42	-0.13	-RP10
	D7S635	−∞	-8.14	-4.89	-2.97	-1.22	-0.49	-0.15	
13	D10S211	−∞	-4.78	-2.03	-0.96	-0.13	0.13	0.14	-GAD2
	D10S199	−∞	-3.37	-1.40	-0.65	-0.07	0.11	0.12	
12	D11S916	−∞	-8.76	-3.63	-1.75	-0.3	0.17	0.21	FTH1
10	D11S901	-2.77	-2.78	-2.71	-2.15	-1.19	-0.63	-0.27	
	D11S1311	−∞	-11.85	-6.65	-4.19	-2.05	- 1	-0.38	
20	D12S345	−∞	-7.42	-4.81	-3.51	-1.76	-0.85	-0.33	PRPH
15	D12S90	−∞	-2.49	-1.16	-0.65	-0.23	-0.06	0.00	
	D12S80	−∞	-1.33	-0.11	0.27	0.42	0.35	0.20	
0	D15S210	−∞	-8.07	-3.8	-2.03	-0.62	-0.09	0.07	-GABRB3
13	D15S122	−∞	-2.11	-0.93	-0.54	-0.25	-0.11	-0.03	
	D15S165	−∞	-3.89	-3.29	-2.99	-1.90	-1.03	-0.44	
22	D19S219	−∞	-8.94	-4.29	-2.44	-0.92	-0.31	-0.06	FTL
8	D19S418	-0.23	2.14	2.49	2.36	1.8	1.18	0.58	
	D19S210	−∞	-3.37	-1.57	-0.85	-0.29	-0.09	-0.02	
9	D22S275	−∞	-8.36	-3.79	-2.08	-0.74	-0.24	-0.05	GAD3
9	D22S277	−∞	-8.17	-4.01	-2.34	-0.96	-0.39	-0.12	
2	D22S423	−∞	-5.75	-2.49	-1.27	-0.34	-0.03	0.05	
	D22S282	−∞	-6.85	-2.93	-1.47	-0.35	0.03	0.10	

Table 3. Exclusion of candidate gene regions. Lod scores are shown for the HS1 family at markers near candidate loci. Genetic distances between markers are listed under "cM" (centimorgans).

Homozygosity mapping (Lander and Botstein, 1987) is based on the principal that homozygous affected individuals whose parents are related most likely have inherited a common haplotype without recombination from a single founder. Homozygosity mapping can be a very efficient, time-saving approach, as one can genotype only affected individuals and look for markers for which these individuals are homozygous. As noted, this method has been successfully used to map several disorders. Homozygosity mapping is most useful in inbred populations, like those found in the Middle East, Japan, and among the American Amish. This method can be used for localization as well as fine mapping of rare recessive disorders. However, to increase our potential to detect linkage, we opted to genotype all individuals in the HS1 family (not just the affected individuals).

Another use of isolated, inbred populations is 'shared segments' mapping (Houwen et al. 1994). This technique makes use of both the population dynamics of these isolated populations and the phenomena of linkage disequilibrium. Disease gene loci are mapped by searching the genome for linkage disequilibrium between the disease and marker loci at approximately 10-cM intervals. This procedure is most applicable when the disease is rare and hence likely to have arisen from only a few founders, as well as when the population is isolated and the disease has arisen within the past few hundred years. Linkage is detected because of the excess sharing of the same chromosomal haplotype of markers among affected individuals within the population, regardless of their degree of relationship. This method can be applied to dominant as well as recessive traits, given the appropriate population and sample size. This method could not be applied in our study because we did not have the stratification of families. Had there been multiple Amish families with HSS then this method might have been an alternative.

Pooling of DNA samples (Sheffield et al., 1994) is another method by which to look for homozygosity and minimize the number of PCR reactions performed. This method involves pooling of DNA from different individuals into a single sample. Pooled DNA from affected individuals is compared to pooled DNA from unaffected siblings and controls. A decrease in the number of alleles in the affected individual pool suggests an increase in homozygosity, as would be expected when examining markers near the disease gene locus. Care must be taken to assure that all individuals are equally represented within the pools. This technique is best suited for tri- and tetranucleotide repeat markers because they are easy to genotype (e.g., few shadow bands). Dinucleotide repeat markers cannot easily be used in pooling strategies because of complicated shadow-banding patterns, which makes interpretation of results difficult. An example of this pooling strategy with several $(CA)_n$ repeat markers is shown in **Figure 3**. There are clearly discrepancies and ambiguities for many of the markers. Alleles found in the affected-children pool are sometimes not detected in the parent pool (i.e., D16S518), and some markers are extremely difficult to read (i.e., D13S285). When this study first started, there were not enough tri- and tetranucleotide repeat markers to evenly cover the genome. However, because we used dinucleotide repeat markers for our genome search, the samples were not pooled.

Another method for conducting a genome search is to run markers in gene-rich regions first. By looking at the location of genes that have been previously mapped, Antonarakis (1994) noted a correlation between gene-rich regions and GC-rich regions of the chromosomes using cytogenetic techniques. However, there were no genetic maps readily available (that could be used for a genome search) that placed highly polymorphic

markers within the gene-rich regions.

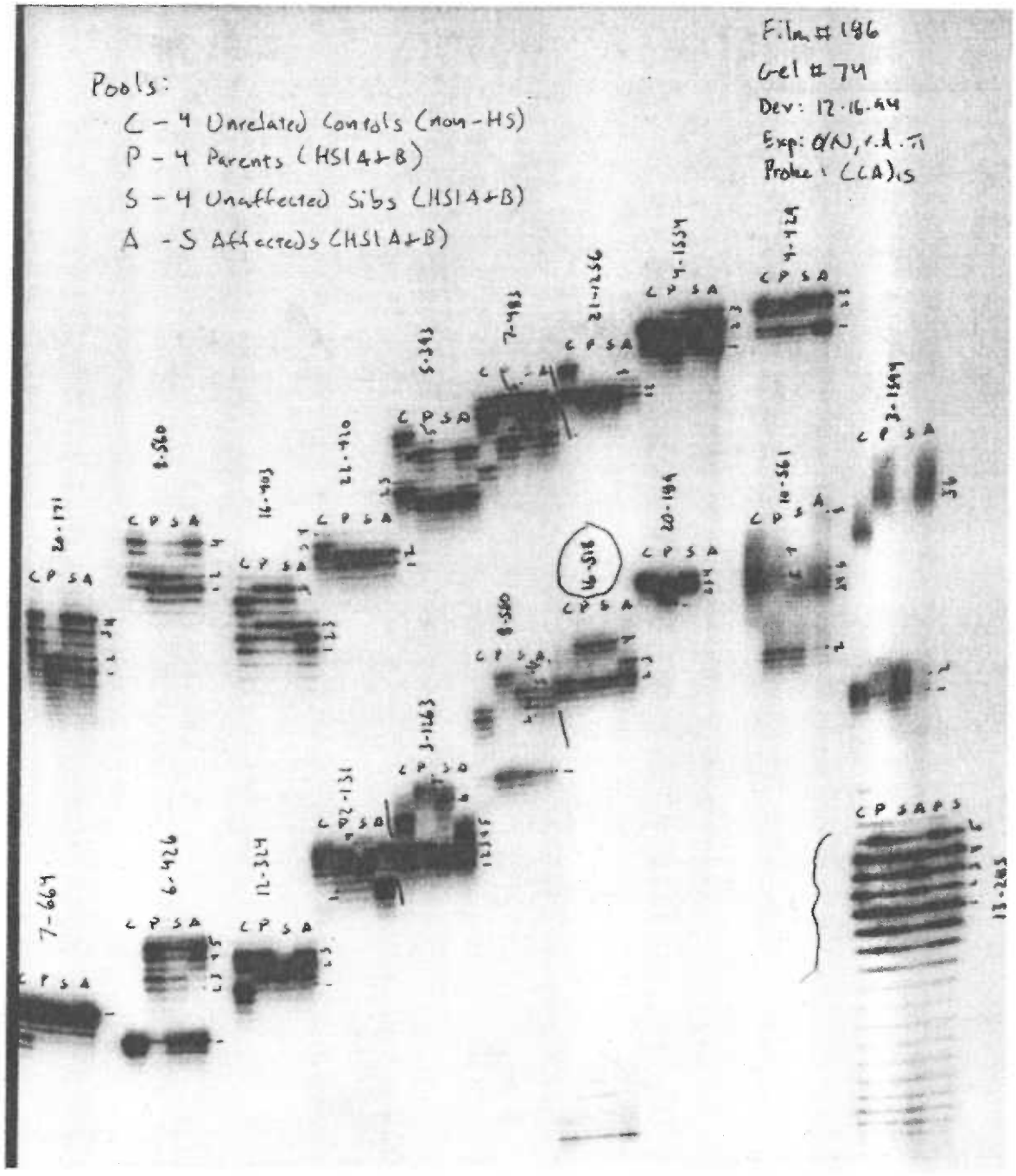


Figure 3. An autoradiograph of PCR products generated using the pooling strategy. Twenty $(CA)_n$ repeat markers were typed using DNA pools from the HS1 family and a set of controls. Each of the four pools contained four different individuals' DNA samples in equal concentrations (the affecteds pool has five samples). The pools are comprised of four groups: affecteds, unaffected sibs, parents, and unrelated controls.

Exclusion mapping (Edwards, 1987) is the demonstration that a polymorphic marker or chromosomal region is not linked to the disease in question. A lod score, z, < -2 at a given distance, θ, from the marker locus is considered evidence for exclusion of the region. Correct marker order and distance are critical for this analysis. Map order information should only be used for markers that have odds of at least 1000:1 over any other possible order. Exclusion of regions can be determined using both two-point and multipoint linkage analysis. Multipoint analysis is generally used to exclude regions between markers on a known map. We demonstrated the use of exclusion mapping in our paper 'Clouston syndrome (hidrotic ectodermal dysplasia) is not linked to keratin gene clusters on chromosomes 12 and 17' (Hayflick et al., 1996; Appendix).

Simulation analysis is usually performed before a study to estimate ideal marker spacing and the probability to detect linkage at the estimated spacing. For the random screening process, we estimated the ideal marker spacing needed to recognize linkage of the HSS gene in the HS1 family by taking into account the number of meioses between the founders and the affecteds. Since there are approximately six generations from the founders to each of the affecteds and this is an autosomal recessive disorder, there are twelve possible meioses (6 generations x 2 meioses per generation). The size of the human genome is roughly 3000 cM. A crossover occurs on average every 100 cM, so there are about 30 crossovers per meiosis (3000 cM / 100 cM). In the HS1 family, approximately 360 crossovers could occur (12 meioses x 30 crossovers per meiosis). The average distance per crossover is 8.3 cM (3000 cM / 360 crossovers). Since it is unknown exactly where these crossovers occur, on average, the maximum distance a selected marker could be from a crossover is 8.3 cM. Therefore, in order to detect all

recombination events (and linkage), the markers should be spaced no more than twice this distance, or 16-17 cM apart (8.3 cM x 2). Using this scheme, the probability with which we could expect to find linkage depends on the allele frequencies (heterozygosities) of the markers used, the distances between the markers, the pedigree structures, and the sample size. While pedigree structure and sample size are known, marker allele frequencies and distances between the markers are not as defined. A program for conditional simulation of marker genotypes, called SLINK (Ott 1989; Weeks 1990), can be used to estimated the power of a linkage study. There are two basic methods of simulations that can be performed: conditional and unconditional. In conditional simulation, markers are simulated as if linked to the disease gene. Conditional simulation studies are useful for determining how often linkage will be detected when a marker is linked to the disease. In unconditional simulation, markers are simulated as if unlinked to the disease gene. An unconditional simulation study is useful for determining how often false evidence of linkage may arise when a marker is unlinked to the disease. Due to the complexity of the HS1 family (number of inbreeding loops and number of individuals), this simulation program could not be run. Some simple simulations were run to estimate the maximum lod score of the HS1 family, as will be described in the next section.

In our first random screen we utilized a set of $(CA)_n$ repeat microsatellite markers. These markers were spaced on average every 40 cM and had an average heterozygosity of about 70%. Genetic distances between the markers were based on published information (Weissenbach et al., 1992; Gyapay et al., 1994; Dib et al., 1996). Since no significant evidence for linkage was found for this set of markers (lod score, z, greater

than 3), we increased the density of the markers to one every 20 cM. The available genetic maps were reviewed and markers that had the highest heterozygosities were selected. Eventually, markers were run between any two markers greater than 16 cM apart. Additional markers were run in regions with suggestive evidence for linkage ($1.5 \leq z < 3$) or in regions where one or more of the markers was uninformative. After running about 400 markers and excluding about 90% of the genome (**Figure 4**), we found significant linkage to marker D20S199 ($z = 5.92$ at $\theta = 0.01$) on chromosome 20p. The average spacing of these markers was 7.5 cM, or about half of the ideal marker spacing we estimated. The reason the markers had to be spaced more closely than we estimated was that not all of the markers were fully informative.

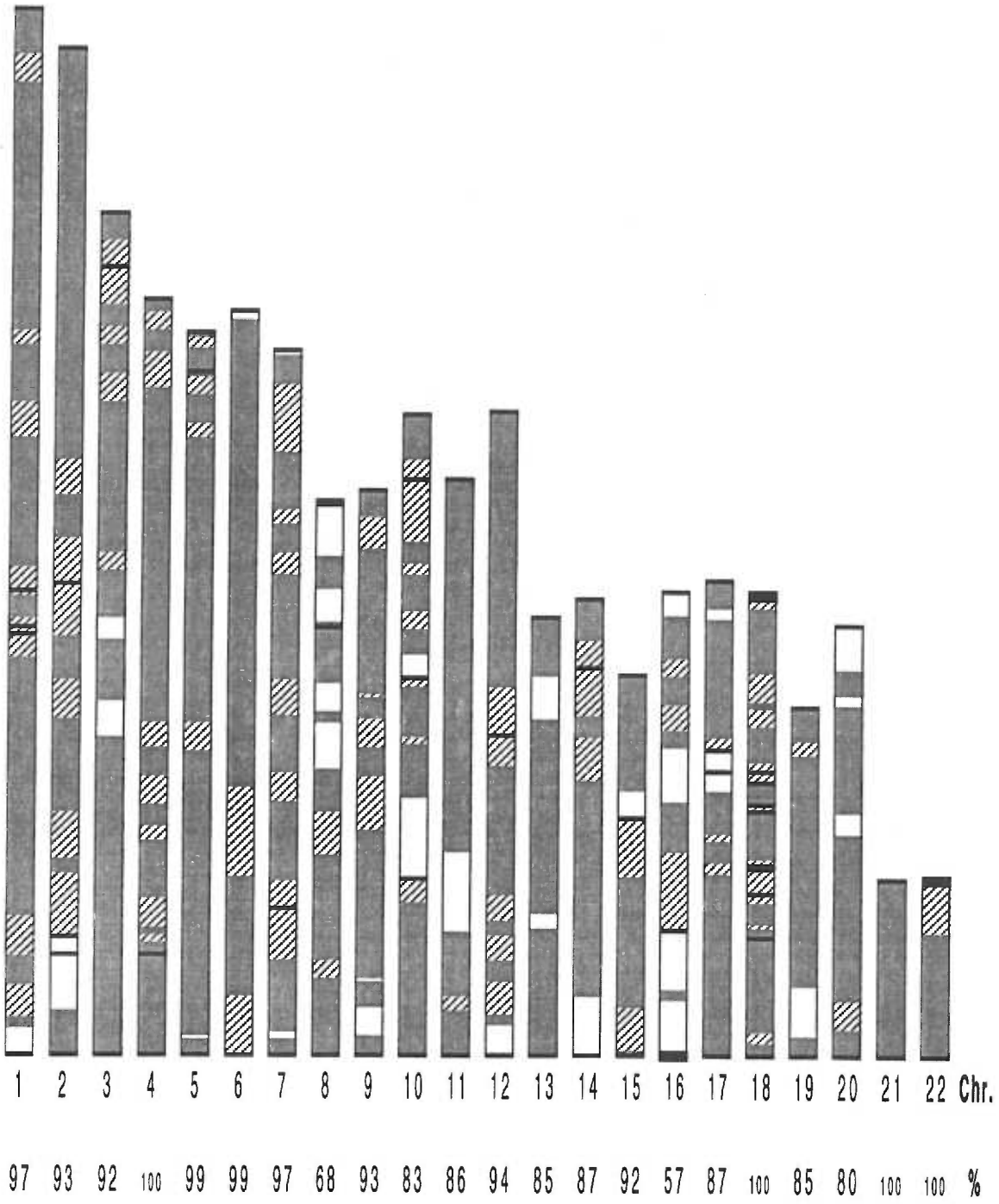


Figure 4. Autosomal regions excluded by the HS1 family. Approximately 90% of the genome was excluded after running 400 markers with the HS1 family. Dark shading indicates regions that were excluded by linkage analysis. Hash marks indicate regions that were excluded by haplotype analysis. Light regions include areas where markers were either uninformative or had not yet been run. Numbers at bottom indicate chromosome number and percent of chromosome excluded, respectively.

LINKAGE ANALYSIS

Linkage analysis is a powerful tool for mapping genetic diseases. Cosegregation of marker and disease genotypes down through a family or set of families is examined to determine if they segregate independently according to Mendel's laws or if they do not segregate independently because of their close physical proximity. Because the underlying biochemical defect remains unknown for most diseases, conventional laboratory mapping methods such as somatic cell hybrid studies and fluorescent in situ hybridization are not applicable. Family studies provide a powerful alternative to these older, laboratory mapping methods. Linkage analysis has proven an extraordinary tool in identification and characterization of disease genes with a simple Mendelian mode of transmission (e.g., Aksentijevich et al., 1993; Kwitek-Black et al., 1993; Hentati et al., 1994).

There are two general approaches to linkage analysis: parametric and nonparametric. Nonparametric methods, which include the affected-sib-pair method (Haseman and Elston, 1972) and the affected-pedigree-member method (Weeks and Lange, 1988), attempt to detect linkage by testing whether the affected individuals have more similarity at a marker locus than expected by chance. Nonparametric approaches do not identify recombinational events directly and so do not provide an estimate of the distance (the recombination fraction) between the marker and disease gene. Parametric methods detect linkage by testing whether the recombination fraction (θ) between the marker and the disease is less than 0.5, where $\theta = 0.5$ indicates that the marker and the disease are unlinked. The main implementation of this approach is the traditional lod score method (Morton, 1955), which attempts to identify and quantify recombination events. This

parametric technique is used for Mendelian disorders, where the mode of inheritance is known and parameters (e.g., gene frequency, type of penetrance, allele frequencies, mutation and recombination rates) can be estimated. When a probable disease-gene model cannot be assumed (e.g., when the genetic etiology is complex) a nonparametric approach should be taken. Nonparametric methods are less powerful than parametric methods, but do not require the specification of the genetic model and other parameters (as parametric methods do). When a specific genetic model can be assumed, as in a Mendelian disorder, a parametric approach should be taken. Incorrect assumptions of the various parameters can severely reduce a parametric method's power to detect linkage, so care must be taken when estimating these parameters.

The lod score method, used in this study, involves several steps. Families segregating the disease gene according to the phenotype are identified. The potential contribution of each family's overall ability to detect linkage, such as the expected lod score, should be measured when possible. Families and individuals that will be genotyped are chosen based on their potential information content. Additional branches of families are collected if needed, and if possible.

A lod score, as first defined by Morton (1955), represents the $\log_{10}$ of the odds for linkage. The likelihood (L) of observing a particular configuration of the disease and marker locus in families is calculated assuming no linkage, or free recombination, between the two loci (when $\theta = 0.5$). This likelihood is then compared with the likelihood of observing the same configuration of the two loci within the same family, assuming varying degrees of linkage over a selected range of recombination fractions (i.e., θ ranges from 0 to 0.5). Log_{10} of the ratio of these likelihoods is then determined for each value of

θ within the range, and each of the resulting numbers is referred to as a lod score, z(x), where x represents a particular value of θ within the range of recombination fractions:

$$z(x) = \log_{10} [L \text{ (pedigree given } \theta = x)/L \text{ (pedigree given } \theta = 0.5)]$$

The value z(x) is referred to as a two-point lod score, as it involves linkage between two loci (the disease gene and a marker). This equation assumes equal recombination for both sexes in the pedigree under consideration. Calculation of lod scores using sex-specific recombination fractions is a simple extension of the above lod score calculation. Differences in recombination fraction by sex have been noted for many chromosomes (NIH/CEPH Collaborative mapping group, 1992) and, when large, should be included in the analysis. The likelihood ratio approach to linkage analysis remains the most powerful linkage method available. In developing this technique, Morton integrated the probability ratio test of Haldane and Smith (1947) with the sequential sampling analysis developed by Wald (1947). Lod score analysis is the most powerful linkage method when mode of inheritance is known, as in the simple Mendelian case. Several software packages are available for calculating two-point lod scores. The one most generally used is the LINKAGE 5.1/FASTLINK 3.0P program package (Lathrop et al, 1984; Cottingham et al, 1993; Schäffer et al., 1994; Terwilliger and Ott, 1994; Schäffer, 1996). This software may be obtained by anonymous FTP from Rockefeller University's Genetic Analysis Software web site (http://linkage.rockefeller.edu/soft/list.html).

If one can assume homogeneity of disease in a given set of families, their lod scores may be summed for each of the corresponding θ values. The overall lod score of the data set is than examined. Lod scores ≥ 3 fall into the significant range. Follow-up studies for

lod scores ≥ 3 include: genotyping of additional markers in the region to confirm the findings, determining variation of marker allele frequencies to show if evidence for linkage is still significant, and blind checking and recording of marker results. A lod score of 3 is equivalent to odds of 10^{-3} (1000:1) in favor of linkage. Lod scores ≥ 1.5 and < 3 generally fall into the promising range. These results indicate that follow-up analyses should be performed for additional markers in the area of interest and for additional families. Once linkage is established or a susceptibility locus is identified and confirmed for the disease in question, fine genetic mapping of the disorder can be pursued. Fine mapping includes genotyping of additional markers in families, as well as the use of additional analyses such as multipoint-linkage or linkage disequilibrium studies. Lod scores between 1.5 and -2 are inconclusive and indicate that additional data are needed. Lod scores of -2 or lower fall into the negative range and indicate regions where it is no longer worth the effort to genotype additional markers (exclusion mapping).

We performed linkage analysis using the LINKAGE 5.1/FASTLINK 3.0P program packages (http://linkage.rockefeller.edu/soft/list.html) on PCs (486DX2-66 MHz and Pentium 233 MHz processors) and on a Sun Microsystems Sparcstation 2 (Palo Alto CA, USA). Most of the calculations shown were done on the 486DX2-66 MHz PC. Some of the basic programs within the package were used for two-point linkage analysis. MAKEPED is a program that converts information on pedigree structure, affection status, and marker genotype from a simple format, generally called the pre-MAKEPED or the pre-file, into a format appropriate for linkage analysis by the linkage programs. PREPLINK is a program that prepares parameter files for linkage analysis by prompting the user for variables specific to the genetic model, such as penetrance and allele

frequencies. LCP, the linkage control program, prepares the batch file that will invoke the analysis programs in their appropriate order. The output from the MAKEPED and PREPLINK programs are necessary input for LCP. The program UNKNOWN performs two main functions and must be run before invoking any of the linkage programs. First, it checks the input pedigree file for data inconsistencies, such as two parents of the same gender or non-Mendelian segregation of marker alleles, and alerts the user if inconsistencies are detected. Second, UNKNOWN looks at a pedigree and deletes any impossible genotypes from consideration in the likelihood calculations for individuals whose disease or marker status is unknown, thereby increasing the speed of the linkage analysis. MLINK is the program that calculates the likelihood of the pedigree under user-specified values of θ for two loci; these likelihoods are used to calculate the lod scores.

For lod score calculations, marker allele frequencies were assumed to be equal (for both sexes) and autosomal recessive inheritance with complete penetrance was assumed using a population gene frequency of 10^{-3}. Markers near candidate gene regions (**Table 2**) were first genotyped and these regions were excluded (**Table 3**) by linkage analysis and haplotype analysis (explained in the next section). Once these regions were excluded, evenly spaced dinucleotide repeat markers were typed across the remaining regions of the autosomal genome. After excluding nearly 90% of the autosomal genome by similar analyses (**Figure 4**), linkage was detected between the HS1 family and marker D20S199 on chromosome 20 ($z = 5.92$ at $\theta = 0.01$). Analysis of nine other families supported linkage of the NBIA1 gene to the chromosome 20p13-p12.3 region with a total maximum two-point lod score of 13.75 at $\theta = 0$ for the marker AFMa049yd1 (see Chapter 2). Linkage analysis was performed independently for each family because of the possibility

of locus heterogeneity. Only one family, HS9, with classical HSS was not linked to this region.

Since the publication of the above results, nearly 60 families have been collected for this study. There are now two families with classical Hallervorden-Spatz syndrome, HS9 and HS50, that are not linked to the chromosome 20p13-p12.3 region (**Table 1, Figure 1**). HS9, a consanguineous family of Japanese descent, was the first classical HSS family found not to be linked to 20p13-p12.3. Homozygosity was not detected at any of the markers from this region, as would be expected for a consanguineous union. HS50, a Saudi Arabian family likely to be consanguineous (because of the high rate of inbreeding in that country), also is not homozygous for any of the markers in 20p. All of the affected individuals in these families show MRI evidence of iron accumulation in the basal ganglia, extrapyramidal dysfunction, and progression of disease. Post-mortem examination of affected members of the two families demonstrated typical findings, including intracellular and extracellular deposition of iron-containing pigment and axonal spheroids in the globus pallidus. The inheritance in these families is consistent with an autosomal recessive pattern. Though the average age of onset in these families is later than in the linked families, at least one affected individual from each unlinked family had onset in the first two decades of life. Furthermore, rate of progression of disease seemed to be slower, with individuals living longer than those in families linked to chromosome 20p13-p12.3. Pigmentary retinopathy was absent in the unlinked families, though one individual did present with optic atrophy. Three clinical features may help to distinguish this group from those with the classical form of this disease. These include later age of onset, slower progression of disease, and absence of retinal pathology. These families

provide evidence for locus heterogeneity in Hallervorden-Spatz syndrome.

The HS1 family was critical for linkage analysis in the early stages of this project. Using this family, we expected to detect homozygosity near the NBIA1 gene region. Because of the large size of the HS1 family and the multiple number of consanguineous unions, there was a great potential to find linkage. However, one problem with having so many consanguineous unions in a kindred is that there are many of what are called 'loops' (six in the case of the HS1 family). There are two types of loops, consanguinity loops and marriage loops, that for linkage purposes are identical. Consanguinity loops occur as result of inbreeding; marriage loops occur when, for example, brothers in one pedigree have children with sisters and another pedigree. Both types of loops lead to what are termed 'complex' pedigree structures. In the linkage programs, loops must be broken by doubling individuals in the pedigree. The way to break a loop is to make one of the participating individuals into two people: one person acts as parent and the other acts as child and sibling. Processing of the pedigree by the linkage utility program MAKEPED will detect loops and alert the user to the presence of undeclared (or unbroken) loops. A loop is said to be present in a pedigree structure when an unbroken line can be drawn, starting from some individual, X, in a pedigree, traversing through connecting relatives, and returning to individual X. without retracing any step. The LINKAGE and FASTLINK programs have a constant called maxloop that, for the programs to run correctly, must be at least as large as the number of loops in any of the input pedigrees. Depending on the data set, the running time may increase exponentially with the number of loops, but the hard disk space requirements will not increase substantially (Schäffer, unpublished). One of the things that the loop encoding system implies is that the user gets

to choose the ordering of the loops. FASTLINK runs on pedigrees with loops can take a long time. When there is more than one loop, each loop should be tried as the first loop, to see which choice runs fastest. By experimenting with the number of loops in the HS1 family (**upper Table 4**; **Chapter 2, Figure 1**), we found that it was only necessary to include three loops in the linkage analysis. Three loops are the minimum number required to connect the founder generation to each of the carrier parents in generation seven. It is evident, from the two-point lod score data in **Table 4**, that at least three loops must be included to retain all of the family information; the lod scores start to decrease when less than three loops are used. However, it is not necessary to include more than three loops in the linkage analysis. When more than three loops are used the lod scores change little, but computation time dramatically increases. In order to reduce the amount of computation time, only three loops were included in the linkage calculations. Changing the order of the loop breakers had no affect on the lod scores and no detectable affect on calculation time (data not shown). Specifying the proband for a family can also have an affect on linkage calculation times; but for this analysis, we opted to let the linkage programs select this parameter.

				Lod score at θ =						
Loops	Alleles	Frequency	Time (min)	0.00	0.01	0.05	0.10	0.20	0.30	0.40
1	5	=	0.03	5.64	5.49	4.91	4.20	2.86	1.65	0.61
2	5	=	0.03	6.63	6.47	5.84	5.06	3.53	2.13	0.92
3	5	=	0.05	7.32	7.15	6.48	5.62	3.92	2.32	0.97
4	5	=	1.5	7.30	7.13	6.46	5.61	3.92	2.34	0.99
5	5	=	31	7.26	7.09	6.43	5.60	3.93	2.35	0.99
6	5	=	11 hr.	7.25	7.08	6.42	5.60	3.93	2.36	1.00
1	10	=	0.03	5.98	5.83	5.22	4.47	3.02	1.71	0.63
2	10	=	0.05	7.21	7.05	6.39	5.56	3.89	2.33	0.99
3	10	=	0.08	8.18	8.00	7.29	6.39	4.54	2.71	1.11
4	10	=	6	8.15	7.97	7.26	6.36	4.52	2.72	1.13
5	10	=	112	8.04	7.87	7.17	6.29	4.50	2.72	1.14
6	10	=	35 hr.	8.00	7.84	7.14	6.27	4.49	2.72	1.14

				Lod score at θ =						
Loops	Alleles	Frequency	Gene Freq.	0.00	0.01	0.05	0.10	0.20	0.30	0.40
3	2	=	0.1	4.85	4.74	4.31	3.75	2.64	1.53	0.51
3	2	=	0.001	6.00	5.88	5.39	4.76	3.48	2.20	0.99
3	2	=	0.00001	6.02	5.90	5.40	4.77	3.49	2.21	0.99
3	2	.1/.9	0.1	5.97	5.83	5.27	4.56	3.12	1.75	0.58
3	2	.1/.9	0.001	7.98	7.83	7.23	6.45	4.81	3.08	1.39
3	2	.1/.9	0.00001	8.01	7.86	7.26	6.48	4.83	3.11	1.41
3	5	=	0.1	5.39	5.25	4.71	4.04	2.72	1.49	0.47
3	5	=	0.001	7.18	7.03	6.43	5.66	4.08	2.52	1.11
3	5	=	0.00001	7.21	7.06	6.45	5.68	4.10	2.54	1.12
3	5	.1/.225	0.1	5.91	5.77	5.18	4.44	2.97	1.61	0.51
3	5	.1/.225	0.001	7.96	7.81	7.18	6.37	4.68	2.95	1.32
3	5	.1/.225	0.00001	7.99	7.84	7.21	6.40	4.71	2.97	1.33
3	5	.01/.2475	0.1	7.18	7.02	6.38	5.54	3.81	2.07	0.63
3	5	.01/.2475	0.001	9.49	9.33	8.67	7.82	6.01	4.05	2.00
3	5	.01/.2475	0.00001	9.52	9.36	8.70	7.85	6.04	4.08	2.02

Table 4. Simulation results. The upper table shows lod scores for the HS1 family using one to six loops for markers with five and ten alleles (with equal frequencies) respectively. Calculation times are approximate and varied from computer to computer. The lower table shows the results of using the HS1 family with 3 loops and various combinations of number of alleles, allele frequencies, and gene frequencies. A slash in the frequency column means that the allele associated with the disease phenotype has the first frequency and the non-associated alleles all have the second frequency.

Autosomal recessive inheritance with complete penetrance was assumed using a population gene frequency of 10^{-3}. Though the gene for HSS is probably much rarer than this, we determined through various simulations that 10^{-3} is a conservative estimate for the gene frequency, and that lowering it only had a slight affect on the likelihood calculations (lower **Table 4**). The HS1 family, with three loops, was used to determine lod scores for diverse combinations of the number of alleles, allele frequencies, and gene frequencies. Specifying a lower gene frequency had little affect on the lod scores. The gene frequency may be higher in the Amish population, but it is rare even in this setting. The HS1 family is the only family to date that has been diagnosed with HSS in the Amish population. We can see from this data that even a marker with only two alleles, when fully informative, can be quite powerful for detecting positive linkage with the HS1 family. Conversely, markers not linked to the NBIA1 gene region, but fully informative, are just as powerful in excluding linkage.

For lod score calculations in the random screening phase of the study, marker allele frequencies were assumed to be equal since reliable population frequencies for many of the markers was not available. Once we had established linkage, we compared the allele frequencies in our families to those in the CEPH databases. Because the allele frequencies in these two groups closely coincided, we then used the published frequencies to recalculate the lod scores (**Chapter 2, Table 1**). The combined pairwise lod scores for HSS families 1, 5, 8, and 10-16, for all linked chromosome 20 markers were calculated using allele frequencies from the CEPH database and compared with the lod scores using equal allele frequencies (**Table 5**). The lod scores were inflated, by an average of 0.59 at $\theta = 0$, when equal allele frequencies were used, except in cases where

the allele associated with the disease phenotype was rare (i.e., markers D20S842 and AFMa049yd1). This tendency for equal allele frequencies to inflate the calculations in favor of linkage when individuals are untyped or when genotypes cannot be uniquely determined, has been shown by Ott (1992). Incorrect specification of allele frequencies can also lead to other incorrect results, including the failure to detect linkage when it is present.

Combined pairwise lod scores for Hallervorden-Spatz syndrome and linked chromosome 20 markers

		Lod scores (Z) at recombination fraction (θ)						
Marker	**Allele Freq.**	**0.00**	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**
D20S113	CEPH	5.70	5.57	5.06	4.39	3.03	1.72	0.60
	equal	7.36	7.20	6.51	5.62	3.79	2.05	0.66
difference		-1.66	-1.63	-1.45	-1.23	-0.76	-0.33	-0.06
GAAT4E12	CEPH	6.20	6.06	5.49	4.75	3.29	1.97	0.88
	equal	5.04	4.92	4.44	3.85	2.69	1.66	0.76
difference		1.16	1.14	1.05	0.90	0.60	0.31	0.12
D20S198	CEPH	9.59	9.36	8.40	7.21	4.91	2.86	1.15
	equal	11.77	11.50	10.40	8.99	6.20	3.58	1.42
difference		-2.18	-2.14	-2.00	-1.78	-1.29	-0.72	-0.27
D20S842	CEPH	13.59	13.29	12.07	10.51	7.38	4.37	1.75
	equal	12.73	12.45	11.27	9.78	6.78	3.93	1.55
difference		0.86	0.84	0.80	0.73	0.60	0.44	0.20
AFMa049yd1	CEPH	13.75	13.46	12.29	10.80	7.78	4.77	1.99
	equal	12.09	11.81	10.65	9.20	6.31	3.60	1.38
difference		1.66	1.65	1.64	1.60	1.47	1.17	0.61
D20S181	CEPH	7.61	7.42	6.68	5.76	3.98	2.38	1.04
	equal	9.09	8.88	8.04	6.98	4.89	2.92	1.25
difference		-1.48	-1.46	-1.36	-1.22	-0.91	-0.54	-0.21
D20S193	CEPH	6.99	6.83	6.14	5.27	3.59	2.10	0.89
	equal	8.66	8.47	7.70	6.69	4.64	2.70	1.11
difference		-1.67	-1.64	-1.56	-1.42	-1.05	-0.60	-0.22
D20S473	CEPH	8.35	8.14	7.29	6.21	4.14	2.35	0.91
	equal	9.73	9.50	8.56	7.35	4.93	2.76	1.04
difference		-1.38	-1.36	-1.27	-1.14	-0.79	-0.41	-0.13
Ave. difference		-0.59	-0.63	-0.57	-0.48	-0.29	-0.11	-0.01

Table 5. Comparison of pairwise lod scores for Hallervorden-Spatz syndrome and linked chromosome 20 markers using the CEPH database allele frequencies and equal allele frequencies. The average difference in lod scores for all markers at each value of theta (θ) is shown at the bottom.

Multipoint linkage analysis, the simultaneous analysis of linkage between a disease locus and several linked markers, is an important step for maximizing linkage information and localizing the disease gene more precisely on established map of markers. Given a fixed, predetermined map of the marker loci used, the disease gene can be positioned along the interval of the map and the most likely position of the disease gene in the mapped region can be determined. For sublocalization of the disease gene, the most likely position is estimated by comparison of the logarithms of likelihoods for the gene being at different positions on the genetic map. Multipoint linkage analysis could not be calculated for the HS1 family due to the number of inbreeding loops. Instead, as will be described, haplotypes were constructed for all families.

While there were many obstacles to overcome in using the HS1 family for linkage analysis, there were ways around them as described above. I completed another project where more traditional linkage analysis was used (see papers in Appendix). This project involved looking at candidate gene regions for Clouston syndrome (CS; hidrotic ectodermal dysplasia), an autosomal dominant trait with nearly complete penetrance. In two publications, we showed CS was not linked to known keratin gene clusters on chromosomes 12 and 17 (Hayflick et al., 1996), and we confirmed linkage of CS to chromosome 13q11-q12.1 (Taylor et al., 1998). There were no inbreeding loops or other caveats in the CS families.

Using approximately 400 polymorphic markers, we excluded HSS from 90% of the autosomal genome. Within this exclusion map were several regions containing candidate genes related to iron transport and metabolism, and loci associated with retinitis pigmentosa. Linkage of HSS to chromosome 20p13-p12.3 was detected using just the

HS1 family, a large Amish kindred, in an initial random screen of the genome. Additional markers and families were genotyped which confirmed this linkage result, with a maximum lod score of 13.75 at $\theta = 0$ for marker AFMa049yd1.

HAPLOTYPE AND LINKAGE DISEQUILIBRIUM ANALYSIS

Haplotype (crossover) analysis involves examining individual crossovers (recombinant individuals) in the family data in order to define the closest distal and proximal markers that flank the disease gene. Haplotypes of the chromosome marker loci are constructed based on established map order in the recombinant individuals. This is done by constructing chromosomal haplotypes for each person involved in the crossover event, including the family member in whom the crossover appears, and his or her parents, grandparents, siblings, and/or children, if necessary. Haplotypes are constructed by examining the ordered markers genotyped in an individual and arranging them by chromatid, based on parental types. These data are used to determine the minimal region of interest. When new markers become available, they are mapped on the region of interest with respect to existing markers. Depending on the results of these analyses-whether or not genotyping of the markers would further narrow the region-they should be examined in the crossover branch of the family. If informative, the haplotype data are then examined to see if the new marker exhibits crossing over. If crossover data are consistent, the area can be further and rapidly refined. Crossover analysis is extremely useful in determining ancestral haplotypes.

Examination of chromosomal haplotyping crossover data in a family serves a two-fold purpose. First, crossover analysis determines the closest proximal and distal markers to a particular disease gene locus. Crossovers from the entire data set can be compiled and depicted graphically (see **Figures 1** and **5**). This graphic representation allows rapid assessment of both overall and individual crossover situations in terms of non-informative regions, missing data, etc. As new markers are developed, they need not be

examined over the entire data set, but only in the key crossover branches. If a new marker maps between unknown flanking markers and is informative, it will help narrow the region of interest. Second, crossover analysis identifies possible errors that would not be identified by two-point or multipoint analysis. For example, if the order of the markers is well known and the markers are tightly linked (i.e., separated by < 20 cM), occurrence of double crossovers is highly unlikely because the region is so narrow. When double crossovers are found in disease pedigrees under these circumstances, they are likely to represent errors and the individuals concerned should be re-genotyped.

There are several software programs (HAPLO, SIMCROSS, and SIMWALK) that can be used for automated haplotyping (Weeks et al., 1995). We found that constructing the haplotypes by hand was not very difficult and opted not to use any of the available programs. In addition, the complexity of the HS1 family prevented us from using some of the programs. By first observing how alleles segregated from parent to child starting at the bottom of each pedigree, and by minimizing the number of crossover events between markers, we could easily construct haplotypes for all the HSS families.

We constructed haplotypes for all families in a 17-cM region spanning the NBIA1 critical region (D20S864 to D20S95, data not shown). Pedigrees (and haplotypes) for the classical HSS families are shown in **Figure 1** (and in **Chapter 2, Figure 1**). For our initial set of markers and families, the NBIA1 critical region was narrowed to a 4-cM region between markers D20S906 and D20S116 (**Chapter 2, Figure 2**). The closest distal and proximal markers to the NBIA1 gene locus were defined by obligate recombinants in HS1 and HS16, respectively. Markers from several maps were genotyped and ordered based on published data (NIH/CEPH Collaborative Mapping

Group, 1992; Weissenbach et al., 1992; Chumakov et al., 1995; Sheffield et al., 1995; Hudson et al., 1995; Gyapay et al., 1996). There were often discrepancies in the ordering of the markers between these maps. A few double-recombinants were found in our haplotype data, and were resolved by either re-genotyping the individuals in question or by altering the order of the markers within the region. Haplotypes were also constructed during the candidate gene region screen and the random genome screen phases. In regions where linkage alone was insufficient to exclude the area between two closely linked markers, haplotype information was often informative enough to exclude the region. When haplotype information could not be used to exclude these regions, at least one additional marker was run between the two flanking markers.

As additional families with HSS became available, they were genotyped and incorporated into the analysis. **Figure 5** shows the updated NBIA1 critical region after including the new, classical HSS families. The NBIA1 critical region was further narrowed by the inclusion of HS50 family. The closest distal marker became D20S473 and the closest proximal marker became D20S867, due to obligate recombinants by the HS50 and HS16 families, respectively. With improved haplotype information, we were able to reorder markers D20S867 and D20S116. Though the HS16 family still defined the proximal end, the closest proximal marker became D20S867. The distance between markers D20S473 and D20S867 is unknown, but it is definitely less than 2 cM (as will be shown). Construction of a physical map will help to resolve the actual distance between these two markers.

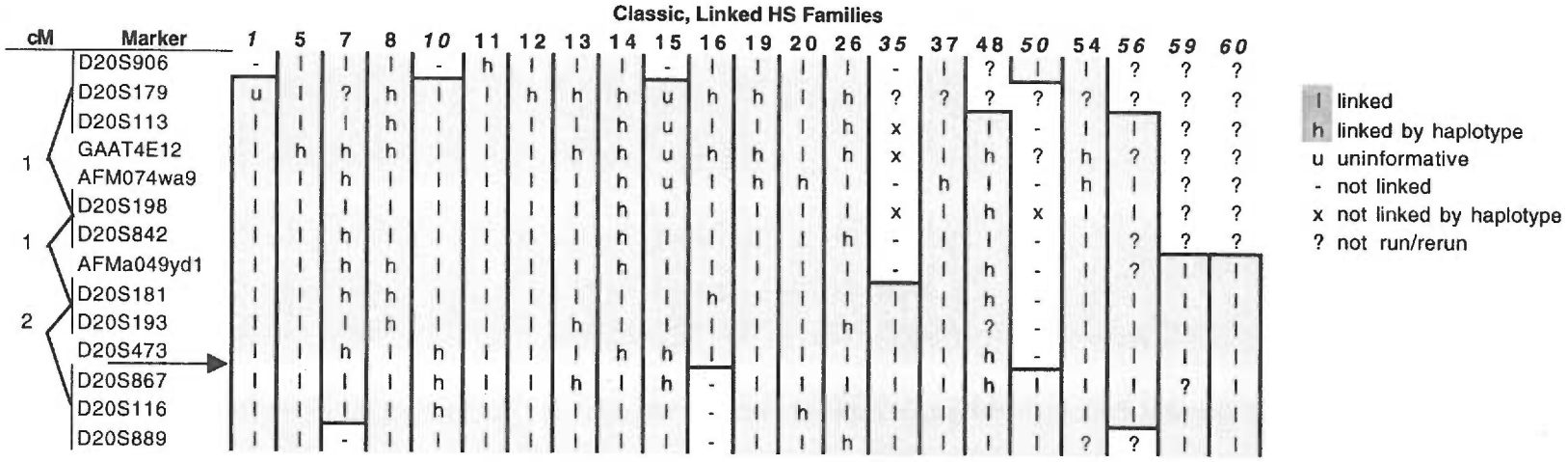


Figure 5. The critical region is defined by recombination events in the classical HSS families. Shaded bars indicate the region in each family that segregates with the disease. Known or likely consanguineous families are indicated by italics. Only the region of homozygosity is considered to be segregating with the gene in consanguineous families (1, 10, 35, 50, 56, 59 and 60). An 'l' indicates linkage by lod score analysis. An 'h' indicates linkage by haplotype. An 'x' indicates exclusion by haplotype. A 'u' indicates an uninformative marker. A '-' indicates obligate recombination with the disease locus. A '?' indicates a marker that has not yet been run.

Because Hallervorden-Spatz syndrome is a rare disorder, it is expected that in inbred families only the regions of homozygosity are segregating with the disease gene. Of the 22 families used in the haplotype analysis, seven are consanguineous (HS1, HS10, HS35, HS50, HS56, HS59 and HS60). These seven families were only considered linked to the NBIA1 critical region for those markers that were homozygous. All of the families, besides HS50, are homozygous at five or more adjacent markers. HS50, a Saudi Arabian kindred, is only homozygous for markers D20S867, D20S116 and D20S889, with respective alleles of 5-3-7 and allele frequencies of 0.018-0.315-0.018. The published genetic distance between these markers is 0 cM, though we did find a few crossovers between them. Based on the high rate of consanguinity in Saudi Arabia, it is very likely that the parents in the HS50 family are related, maybe only distantly. The allele frequencies listed here are for the CEPH families and may not correctly represent the Saudi Arabian population. The probability that the parents in HS50 share these three alleles given that they are not related to each other is simply the product of the allele frequencies:

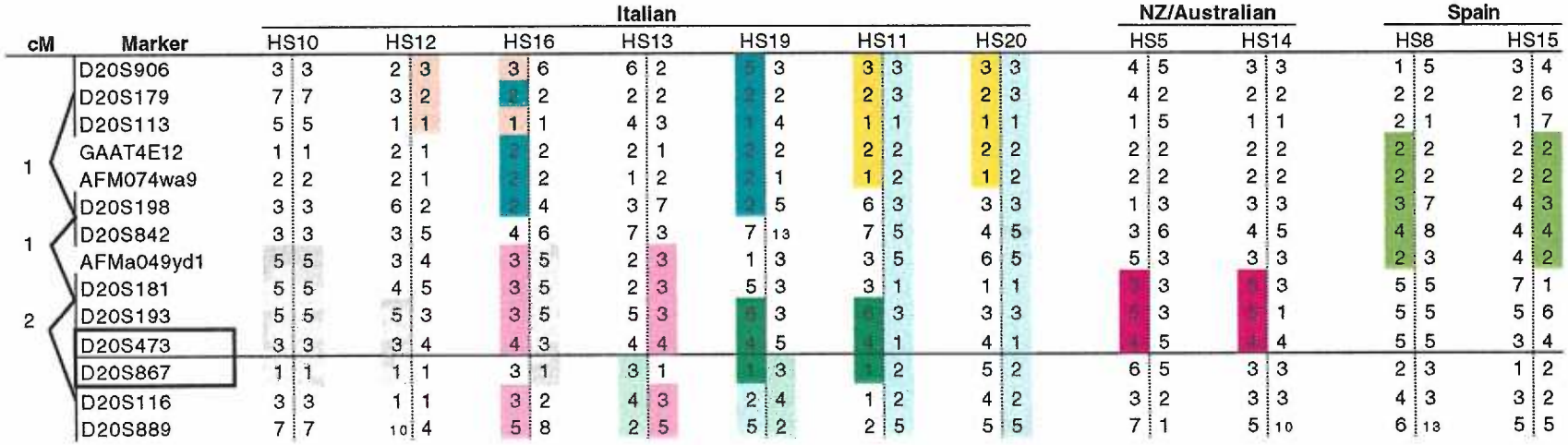
$$P(\text{share 3 alleles} \mid \text{not related}) = 0.018 \times 0.315 \times 0.018 = 0.0001 = 0.01\%$$

Given this small percent and the fact that these markers are very closely linked to one another, it is very unlikely that the parents in HS50 are not related. They may be sharing several other regions of the genome as well, but it is very coincidental that they happen to be sharing haplotypes in the NBIA1 critical region.

Several of the HSS families (i.e., HS8 and HS12) are homozygous for only a few markers up to, but not spanning, the D20S473-D20S867 critical region. It is possible that

the parents in these families are very distantly related, and that there is a small region of homozygosity which has yet to be identified between these two markers in these families. Alternatively, the parents may just happen to be homozygous in this region by chance. Additional polymorphic markers will be developed in this region and they will be used to genotype any of the families that may provide additional crossover (i.e., HS16) or homozygosity (i.e., HS50) information.

Few of the classical HSS families are of the same ethnic descent. As stated earlier, there are three families from Canada, three families from Spain, four families from the USA, and seven families from Italy. There are also two families from the New Zealand/Australia region. After haplotypes were constructed, we looked at several of these families and noted any regions where they were sharing three or more adjacent alleles (**Figure 6**). The families from Canada and the USA were of varying ethnic descent so were not included in this analysis. The third Spanish family, HS48, was not available at the time. The regional distribution for the remaining families is not known. Although all of the Italian families were ascertained through one institute, their geographical distribution is unknown. By highlighting the various shared haplotype segments in the original 4-cM critical interval with different colors, it was observed that all of the families, except for the two Spanish families, shared at least one chromosomal haplotype with at least one other family. These shared segments either flanked or spanned the NBIA1 critical region between markers D20S473 and D20S867. There was no single, common haplotype, but several different ones of varying length. No statistical measure of this sharing was calculated. As might be expected in a linkage disequilibrium study, there was no common allele being shared by all of the families in this region. As demonstrated



cM	Marker	Italian												NZ/Australian				Spain			
		HS10		HS12		HS16		HS13		HS19		HS11		HS20		HS5		HS14		HS8		HS15	
	D20S906	3	3	2	3	3	6	6	2	5	3	3	3	3	3	4	5	3	3	1	5	3	4
	D20S179	7	7	3	2	2	2	2	2	2	2	2	3	2	3	4	2	2	2	2	2	2	6
	D20S113	5	5	1	1	1	1	4	3	1	4	1	1	1	1	1	5	1	1	2	1	1	7
	GAAT4E12	1	1	2	1	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	AFM074wa9	2	2	2	1	2	2	1	2	2	1	1	2	1	2	2	2	2	2	2	2	2	2
	D20S198	3	3	6	2	2	4	3	7	2	5	6	3	3	3	1	3	3	3	3	7	4	3
1	D20S842	3	3	3	5	4	6	7	3	7	13	7	5	4	5	3	6	4	5	4	8	4	4
	AFMa049yd1	5	5	3	4	3	5	2	3	1	3	3	5	6	5	5	3	3	3	2	3	4	2
	D20S181	5	5	4	5	3	5	2	3	5	3	3	1	1	1	3	3	3	3	5	5	7	1
2	D20S193	5	5	5	3	3	5	5	3	3	3	3	3	3	3	3	3	1	1	5	5	5	6
	D20S473	3	3	3	4	4	3	4	4	4	5	4	1	4	1	5	5	4	4	5	5	3	4
	D20S867	1	1	1	1	3	1	3	1	1	3	1	2	5	2	6	5	3	3	2	3	1	2
	D20S116	3	3	1	1	3	2	4	3	2	4	1	2	4	2	3	2	3	3	4	3	3	2
	D20S889	7	7	10	4	5	8	2	5	5	2	2	5	5	5	7	1	5	10	6	13	5	5

Figure 6. Haplotype sharing within ethnic groups in families with classical HSS. Shared haplotypes are indicated by similar coloring. The boxed markers (D20S473 & D20S867) and horizontal line indicate the NBIA1 critical region.

Another method used in fine mapping, as an adjunct to standard linkage analysis, is linkage disequilibrium analysis. Linkage disequilibrium is the condition where two alleles are found together in a population at a greater frequency than that predicted simply by the product of their individual allele frequencies. Along with haplotype analysis based on crossover studies, disequilibrium analysis can be a powerful tool in defining the region of interest. Such studies are usually undertaken when the limit of resolution allowable by crossover and linkage analysis has been reached. Many different measures of linkage disequilibrium have been proposed (Hill and Robertson, 1968; Chakravarti et al., 1984). Linkage disequilibrium studies are well suited to both dominant and recessive disorders that do not appear to exhibit genetic heterogeneity and that appear to have a low mutation rate. Disequilibrium studies have also proven especially useful in mapping disorders that arise from specific isolated populations in which a founder is evident (Hästbacka et al., 1992; Ozelius et al., 1992).

Linkage disequilibrium occurs when a particular marker allele is associated with a disease trait locus at a greater than expected frequency across multiple families. Association of a marker locus with a disease locus is a population-based phenomena. However, when a marker and disease loci are very close together on a chromosome, genetic crossing over will have occurred at such a low rate that the marker will cosegregate with the gene even over many generations. Thus, if a single mutation is responsible for the majority of cases in a population (founder affect), then the same marker allele will also show up in these affected individuals regardless of the family studied. This is in contrast to situations where the two loci are further apart but still linked, in which case repeated crossing over will allow all possible combinations of

chromosomal haplotypes to appear with frequencies as predicted by the equation for Hardy-Weinberg equilibrium (Hardy and Weinberg, 1908). Thus, linkage disequilibrium can be very useful in defining the ancestral haplotype of a disease gene in relation to several marker loci. It can be used for fine mapping of the disease gene even when complete linkage ($\theta = 0$) is established in the families being studied. In general, the stronger the disequilibrium, the closer the marker is to the disease locus. This is not always the case, however, because the level of observed disequilibrium is also affected by allele frequencies and mutation rates at the marker locus.

Several software programs are available that use the maximum likelihood (ML) approach to evaluate linkage disequilibrium. Software for the ML approach of Terwilliger (1995) was obtained by anonymous ftp at ftp.well.ox.ac.uk. In this package, several programs can be used to test for linkage disequilibrium. The programs DISLAMB, which evaluates single locus linkage disequilibrium analysis for a multiallelic system, and DISMULT, which performs multilocus linkage disequilibrium analysis, were used. Each family can only contribute four alleles to the analysis: two case alleles (from the first affected offspring) and two control alleles (the non-transmitted parental alleles). The results of the DISLAMB analysis are shown in **Table 6**. No significant results were obtained with either analysis (the results for DISMULT are not shown, as all calculated lod scores were either zero or very close to zero), most likely due to the small sample size. In order to obtain significant results with linkage disequilibrium analysis, a large number of families are required. In contrast, we had only seven Italian families in this study. In some cases, for highly polymorphic markers, the analysis could not be run because there were too few of any given allele.

Marker	LRT Chi-Square	p-value	Lambda	Significant?	2 x n table Chi-square	P-value	Significant?
D20S113	0.000	0.500	0.000	No	0.755	0.860	No
GAAT4E12	0.000	0.500	0.000	No	0.011	0.918	No
D20S198	0.000	0.500	0.000	No	4.133	0.530	No
D20S842	0.000	0.500	0.000	No	5.598	0.347	No
AFMa049yd1	0.000	0.500	0.000	No	5.578	0.233	No
D20S181	0.000	0.500	0.000	No	0.535	0.911	No
D20S193	0.294	0.294	0.354	No	7.907	0.048	No
D20S473	0.000	0.500	0.000	No	1.810	0.613	No

Table 6. Linkage disequilibrium analysis using the DISLAMB program for seven Italian families. Disease allele frequency equals 0.001 (3 $\log^{-10}$ units for support interval construction).

Once a gene has been mapped to a region, that region can then be examined for any cloned genes that may have been previously mapped there. These genes may be considered as candidates for the disorder if there is some biochemical, physiological, or pathological reason to do so. Fine mapping of the region and crossover analysis can help to minimize the critical gene region, and subsequently, the number of candidate genes to be examined in that region. Correlation of the region with that in other species, such as mouse, can be made by looking at the syntenic regions in those species. The syntenic region to human chromosome 20p13-p12.3 is on mouse chromosome 2, at 75 cM. Gene order in this region of the mouse is relatively conserved with that in humans. There are currently no additional cloned genes in this region of the mouse that make for tantalizing candidate genes for Hallervorden-Spatz syndrome.

In the initial 4-cM NBIA1 critical region, there were several previously mapped cloned genes. These genes included prodynorphin (PDYN; Litt et al., 1988; Summar et al., 1990), adrenergic alpha 1D receptor (ADRA1D; Yang-Feng, 1994; Esbenshade et al., 1995), phospholipase C, beta-4 (PLCB4; Alvarez, 1995), oxytocin (OXT; Summar et al., 1990; Gopal et al., 1992), FK506 binding protein (FKBP1; DiLella, 1992; Lam, 1995), protein-tyrosine phophatase (PTPRA; Jirik, 1992), and sialoadhesin (SN; Mucklow, 1995; Kelm, 1996). Many uncharacterized expressed sequence tags (ESTs) have also been mapped to the NBIA1 critical region. Until we had better mapping information, prodynorphin was considered a potential candidate gene. It is expressed in the basal ganglia and it encodes a neuropeptide precursor that is also present in regions of high iron accumulation in HSS patients. Some mutation analysis (not described here) was performed for PDYN with several HSS patients using the Non-Isotopic RNase Cleavage

Assay (NIRCA; Goldrick et al., 1996) but with no significant findings. When the mapping information in our region improved and we were able to further narrow the NBIA1 critical region, the PDYN gene was no longer within the candidate region.

One candidate gene that has not been excluded from the critical region between D20S473 and D20S867 is sialoadhesin. SN is a cell-cell interaction molecule expressed by subpopulations of tissue macrophages (Mucklow et al., 1995). It contains 17 immunoglobulin-like domains and is structurally related to three other members of the sialoadhesin family. Studies in mice implicate that the mouse homolog (Sn) may play a role in the development of myeloid cells in bone marrow and in the trafficking of leukocytes in lymphatic organs (Kelm et al., 1997). Interestingly, one of the other members of the sialoadhesin family, myelin-associated glycoprotein (MAG), is expressed in myelin of oligodendrocytes and Schwann cells and specifically targets neurons and oligodendrocytes. It may be that in humans, SN interacts with or modulates synaptic transmissions in a manner which when disrupted leads to a cascade of events like those in the pathophysiology of HSS. Preparations are underway to screen several of the HSS individuals by Southern and Northern blot analysis using a cDNA probe for sialoadhesin.

For the initial fine mapping of the NBIA1 gene critical region, we tried to employ a number of techniques: haplotype (crossover) analysis, shared haplotype identification, and linkage disequilibrium. Some of these techniques were more successful than others in helping to refine the map. The reasons for the differences in success mainly depended on our family materials. There were too few families from any single ethnic background or genetic isolate to perform meaningful linkage disequilibrium results. Identifying shared haplotypes did not help to narrow the NBIA1 region, but gave strong evidence that some

of the families are distantly related, perhaps unbeknownst to the families in question. By identifying shared chromosomal segments within the Italian families, there was suggestive evidence that the NBIA1 region is between markers D20S473 and D20S867. Construction of haplotypes, in conjunction with genotyping of additional families and markers, provided the most significant data. By including additional families, especially HS50, we were able to reduce the NBIA1 critical region to between markers D20S473 and D20S867, which are within 2 cM (according to the CEPH map). Additional families that had obligate recombinants or were homozygous also helped to strengthen the refined NBIA1 region. When double-recombinants occurred, individuals were re-genotyped and markers were sometimes reordered to remove them. Mapping of the NBIA1 gene to chromosome 20p13-p12.3 introduced us to several candidate genes, and fine mapping of the region helped us to exclude many of them.

RADIATION HYBRID MAPPING

Since many of the candidate genes on chromosome 20p13-p12.3 could not be included or excluded from the NBIA1 critical region solely based on the haplotype data, we began to construct a physical map for the region. No high-resolution map of chromosome 20p13-p12.3 was available that integrated all types of markers. Diverse, publicly available maps included: genetic linkage maps (Weissenbach et al., 1992; Gyapay et al., 1994; Murray et al., 1994; Utah Marker Development Group 1995; Dib et al., 1996), radiation hybrid maps (Cox et al., 1990; Hudson et al., 1995; Schuler et al., 1996), and maps based on overlapping YACs (Cohen et al., 1993; Chumakov et al., 1995; Hudson et al., 1995). The correlation between these maps was rather poor in the NBIA1 critical region. There were many inconsistencies in marker order, and not all available markers were on every map. A high-resolution radiation hybrid panel was first used to integrate markers in this region from these maps. Radiation hybrid (RH) maps are based on the analysis of maker-locus retention or loss in a panel of RH clones, are dependent on ease or difficulty of marker typing, and rely on even retention of donor DNA into the host cells. Because radiation hybrid mapping is statistical, the data do not suggest a unique order but rather a most likely order. This method can be quickly incorporated by a lab because the only techniques required are PCR and the ability to run agarose gels. Radiation hybrid mapping is not very expensive and is any easy way to start consolidating markers from different maps, since it relies on simple plus or minus screening.

Radiation hybrid mapping is a somatic cell hybrid technique that was developed to construct high-resolution, contiguous maps of mammalian chromosomes. This method

requires screening a panel of donor/recipient hybrid clones with DNA markers by PCR, and analyzing the retention patterns of the donor DNA. Markers that are physically close should have similar retention patterns. RH mapping provides a method for ordering DNA markers spanning millions of basepairs of DNA at a resolution not easily obtained by other mapping methods. Some of the advantages of RH mapping are (1) distance estimated by this method is directly proportional to physical distance, (2) nonpolymorphic DNA markers, that can not be used for meiotic mapping, can be used for this method, and (3) a high-resolution map that is not easily made by other methods can be obtained.

Building on the earlier work of Goss and Harris (1975, 1977ab), Cox and his colleagues (1990) have demonstrated that radiation hybrid mapping provides a powerful method for fine-structure mapping of human chromosomes. Cox et al. used the method of moments and the analysis of two and four loci at a time to estimate distances between loci and to determine locus order. In contrast, Boehnke et al. (1991) and Lange et al. (1995) have developed multipoint-mapping methods that make use of information on many loci simultaneously. These methods are based on (1) minimizing obligate chromosome breaks, and (2) maximizing the likelihood for several different breakage and retention models.

RHMAP version 3 (Boehnke et al., 1991; Boehnke, 1992; e-mail Dr. Michael Boehnke at boehnke@umich.edu; http://www.sph.umich.edu/group/statgen/software) is a set of three FORTRAN 77 programs that provides the means for a complete statistical analysis of RH mapping data. RH2PT is a program for data description and two-point analysis. RHMINBRK is a program for multilocus ordering by minimization of the

number of obligate chromosome breaks. RHMAXLIK is a program for multilocus ordering by maximization of the likelihood of the hybrid data under a variety of breakage and retention models. These last two programs can evaluate a user-specified list of locus orders, or can employ one of several strategies of combinatorial optimization to attempt to identify the best locus orders. Both of the multipoint methods can be used to identify influential hybrids that have a large impact on ordering conclusions.

RH2PT provides estimates of locus-specific retention probabilities and pairwise breakage probabilities, two-point lod scores for linkage of the various marker pairs, and linkage groups. Linkage groups are indicated for loci that are linked on the basis of two-locus lod scores of at least 4.0, at least 6.0, or at least 8.0 (these were originally set at 2.0, 3.0 and 4.0 when our analyses were performed). It is assumed that (1) breakage is at random along the chromosome, with constant intensity and no interference (in probabilistic terms, breakage along the chromosome is a Poisson process); (2) different chromosomal fragments are retained independently in the resulting RHs; and (3) retention probabilities for the various fragments are all equal.

RHMINBRK calculates the number of obligate chromosome breaks for locus orders, and attempts to identify those orders requiring the fewest obligate chromosome breaks (Boehnke et al., 1991; Bishop and Crockford, 1992; Boehnke, 1992; Weeks et al., 1992). The idea behind the minimum break approach is that two loci lying close together on the chromosome should have fewer breaks between them than two loci lying far apart. Thus, the best locus order is that requiring the fewest obligate chromosome breaks. Such an approach is analogous to genetic mapping by minimizing recombinants (Thompson, 1987). The minimum obligate breaks approach requires only that loci be arranged in a

linear way along the chromosome. Thus, minimum obligate chromosome breaks provides a non-parametric method for locus ordering. Counting obligate breaks is straightforward. For a given locus order, obligate breaks occur when a retained locus follows a locus which is lost or vice versa; untyped loci are ignored. The number of obligate breaks is generally substantially less than the number of actual breaks. If r is the probability a human chromosome fragment is retained in a hybrid, the mean values of the number of obligate breaks (B) and number of actual breaks (N) are related according to $E(B) = 2r(1-r)E(N)$ for haploid radiation hybrids (Barrett, 1992). Thus, for haploid hybrids, the number of actual chromosome breaks will on average be at least twice as large as the number of obligate breaks.

Proper analysis of RH mapping data involves several steps. RHMAP provides a set of programs to carry out many of these steps. The following approach to the analysis of RH mapping data is suggested: (1) careful marker scoring, data entry, and data checking; (2) calculation of descriptive statistics and two-point analysis by RH2PT; (3) nonparametric multipoint RH mapping by RHMINBRK; and (4) maximum likelihood RH mapping by RHMAXLIK. Scoring of markers in RH mapping studies is not trivial, particularly because PCR is used. In cases of ambiguity, such as weak positives, consideration should be given to repeating the experiment or simply calling the data missing. When only a small portion of the data is ambiguous, treating that data as missing generally does not result in substantial information loss.

Once data have been entered, RH2PT can be used to determine linkage groups, estimate locus retention probabilities, and estimate distances between the various markers. Subsequent multipoint analyses (see below) are best undertaken on those (sets

of) loci that appear to be linked based on the two-point analyses; inclusion of unlinked markers complicates the analysis and interpretation of the RH mapping data. Retention probability estimates for the different loci will suggest whether the equal fragment retention probability model should be appropriate for the maximum likelihood analysis, or whether alternative fragment retention models should be considered. Breakage probability and distance estimates for the marker pairs will suggest locus orders for the markers; these orders can be compared to those inferred in the more complex multipoint analyses. Substantial discrepancies suggest checking the analyses undertaken. Breakage probability estimates for the two-point and multipoint maximum likelihood analyses generally should be similar, particularly for loci that are close together. The estimates from these analyses should be compared for consistency. Substantial discrepancies again suggest checking the analyses undertaken.

We incorporated the method of RH mapping as a start to the physical-mapping phase of this project. The method for typing the RH panel is described in the Protocols section. An example of RH screening is shown in **Figure 7**. We used the high-resolution Stanford TNG3 panel created at the Stanford Human Genome Center in order to integrate markers from numerous maps of chromosome 20. This panel consists of 90 radiation hybrid clones of the whole human genome, plus two control DNAs (RM donor and A23 recipient). The human lymphoblastoid cell line (donor RM) was exposed to 50,000 rad of X-rays and fused with nonirradiated thymidine-kinase deficient hamster recipients cells (A23), creating a panel of independent somatic hybrid clones with 60 kilobase (kb) resolution.

For RH mapping, the clones are aliquoted into microtiter plates (see Worksheet 1 for

configuration) that are then used to set up PCR reactions. PCR was performed almost identical to that described earlier, but for this analysis it was sometimes better to optimize the conditions for each marker. The highest annealing temperature that generated robust PCR was used to produce gels that were more easily readable. A 4-5 minute hold at 94°C was also incorporated for all reactions in order to maximize denaturation of the DNA. Agarose gels were poured using six rows of 18-tooth combs. This created enough wells such that an entire 96-well PCR plate could be run at one time. The clones were arranged in the microtiter plates in a staggered formation so that they could be loaded with an automated multi-channel pipette (0.2-10 µl, Rainin Instrument Company, Woburn MA, USA). The agarose gels could be loaded quickly (to prevent diffusion), and the clones remained in numerical order (the multi-channel pipette loads every other well).

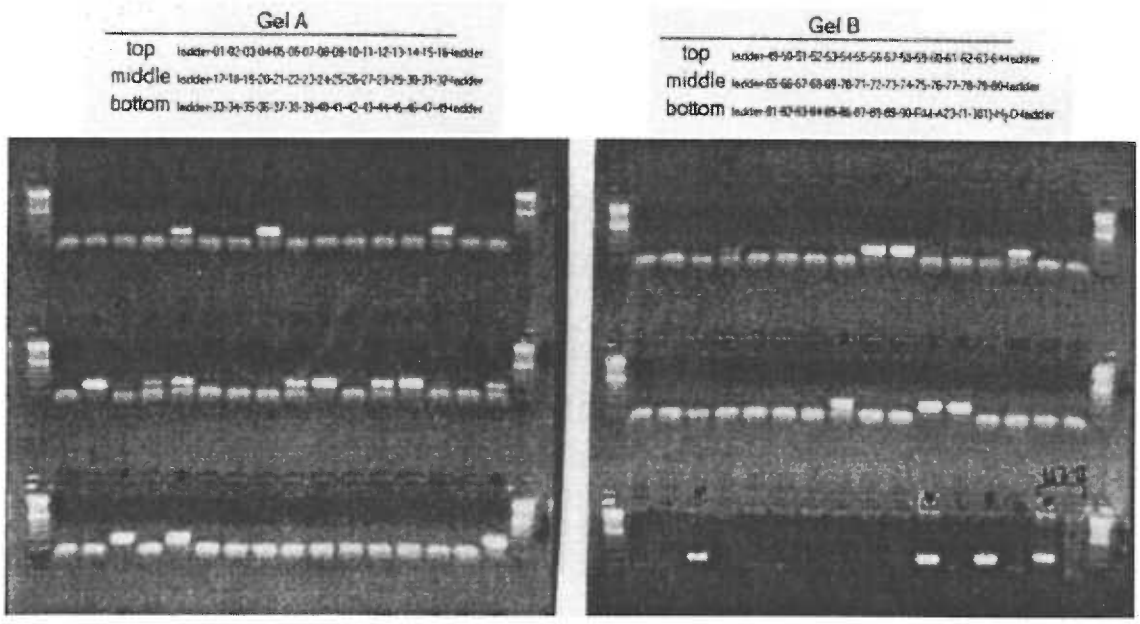


Figure 7. An ethidium bromide (EtBr) stained agarose gel of PCR products generated from the Stanford TNG3 High Resolution Radiation Hybrid Panel using marker D20S482. The last six lanes of Gel B, bottom row, correspond to human control RM, hamster control A23, random human genomic DNA, H2O control, and two different test reactions with another marker. A 100-bp ladder was run on all ends.

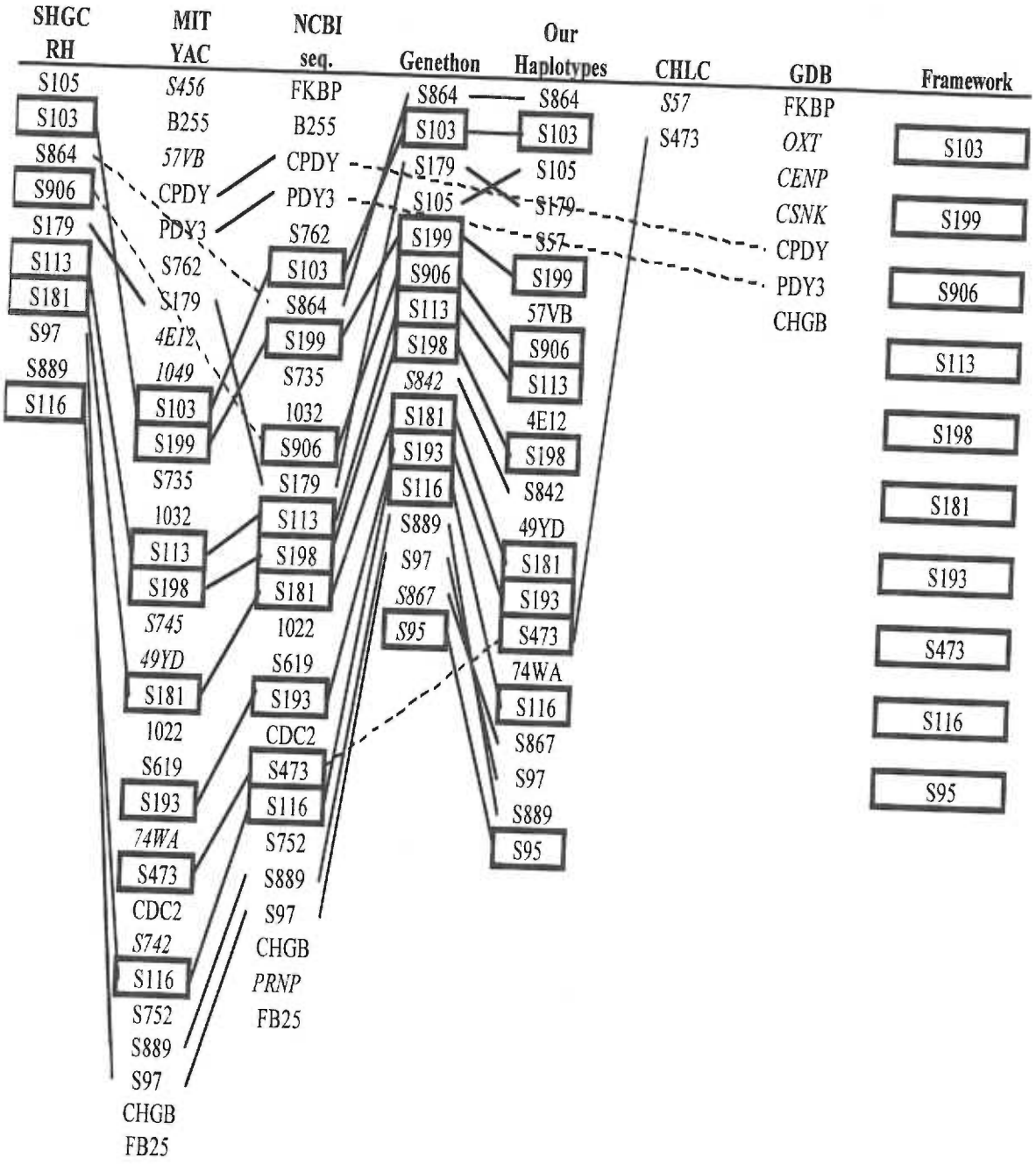


Figure 8. The NCBI integrated map used to determine framework markers for radiation hybrid screening analysis. Map sources are indicated at the top. Framework markers (boxed) were consistently ordered on all maps on which they appeared. Lines connect markers that are on more than one map.

Figure 8 shows the National Center for Biotechnology Information (NCBI) integrated map that was available when we started the physical mapping phase. There is clearly poor correlation between the maps. There are many inconsistencies as to marker ordering, few markers are on all of the maps, and none of the maps is sufficiently marker dense. Markers that were consistently ordered on all of the maps, on which they appeared, were used as framework markers (**Figures 8-10**). This framework set of markers was used to help analyze the RH mapping data.

Analysis was performed with the programs RH2PT and RHMINBRK from the RHMAP software package (Boehnke et al., 1991; Boehnke, 1992). RH2PT was first used to assign 37 markers to linkage groups (**Figure 9**), where each linkage group was assigned a minimum lod score cutoff. The higher the lod score the more stringent the ordering. A lod score here is equivalent to the typical lod score used in linkage mapping (i.e., a lod score of 3 corresponds to 1000:1 odds in favor of linkage). As the lod score cutoff increased, the markers fell into several, smaller linkage groups. The RHMINBRK program was then used to order the markers within each linkage group. Getting the RHMAXLIK program to run was not trivial, and we had already started to screen the BAC DNA libraries, so this analysis was not used on our data. The results of RH2PT plus RHMINBRK are shown in **Figure 10**. The ordering of the linkage groups was determined (1) by framework markers within the groups and (2) by yeast artificial chromosome clone/sequence-tagged site (YAC/STS)-content mapping data generated with the markers (**Chapter 3**). Ordering of markers within linkage groups was determined by the RHMINBRK program (data not shown). The resultant marker order is consistent with that subsequently determined in the BAC screening analysis. This

radiation hybrid mapping data proved useful as an initial fine-mapping tool to quickly integrate markers from several different maps.

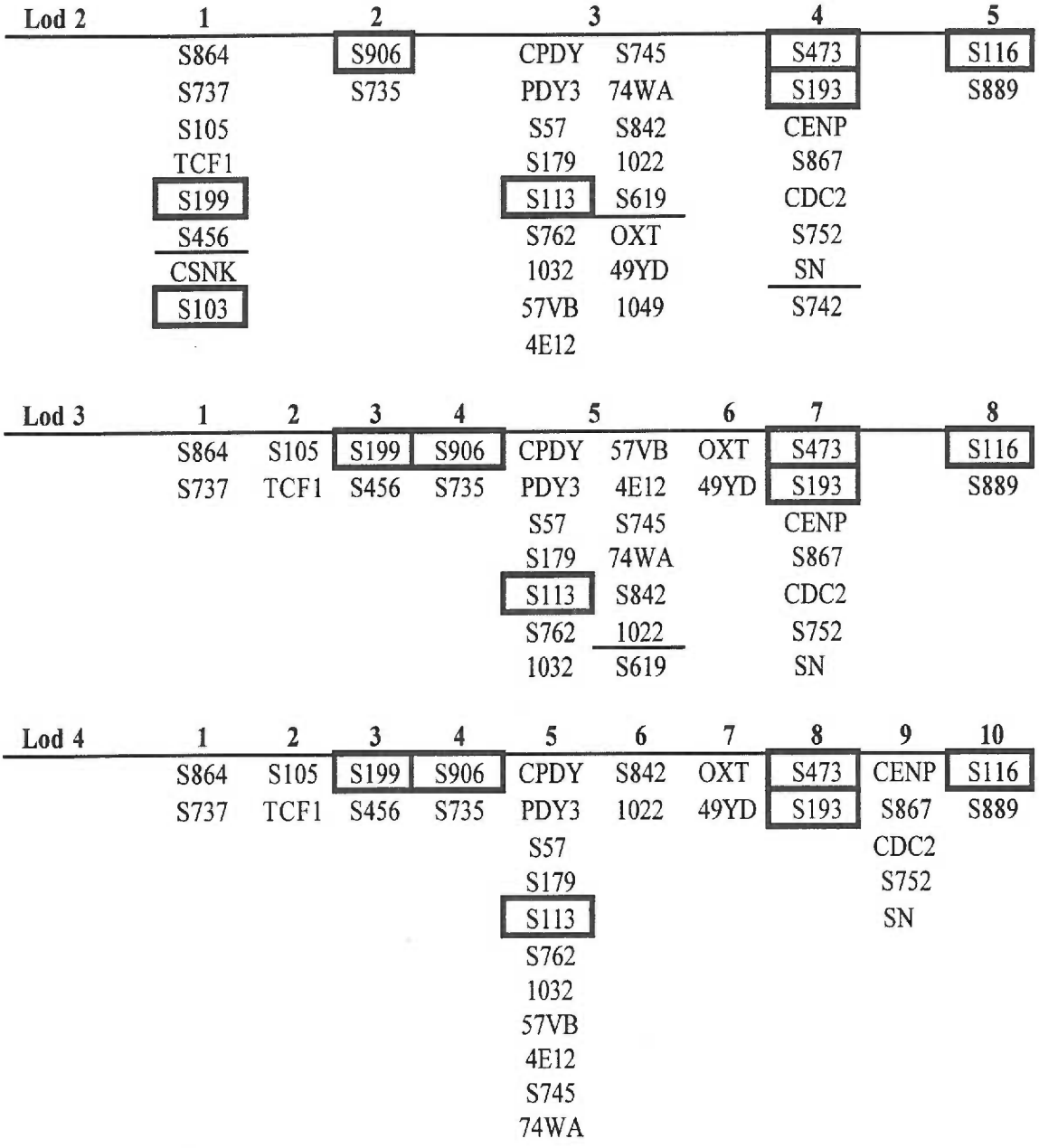


Figure 9. Marker linkage groups from two-point radiation hybrid analysis. Linkage groups are listed in increasing order of significance (i.e., lod 3 corresponds to 1000:1 odds in favor of linkage). Markers in the same linkage groups at the higher lod scores are more likely to be physically close to one another. Framework markers are boxed.

Figure 10. Radiation hybrid mapping data. Markers are indicated at top and are ordered according to two-point and minumum break analyses. Framework markers are boxed. Radiation hybrid clones are indicated at left. Shading indicates strong positive PCR band. Minus signs indicate no PCR band was detected. Question marks indicate undeterminable PCR result.

CHAPTER 2

HOMOZYGOSITY MAPPING OF HALLERVORDEN-SPATZ SYNDROME TO CHROMOSOME 20p13-p12.3

Homozygosity mapping of Hallervorden-Spatz syndrome to chromosome 20p12.3–p13

Todd D. Taylor[1], Michael Litt[1], Patricia Kramer[1,2], Massimo Pandolfo[3], Lucia Angelini[4], Nardo Nardocci[4], Suzanne Davis[5], M. Pineda[6], Haruo Hattori[7], Peter J. Flett[8], M. Roberta Cilio[9], Enrico Bertini[9] & Susan J. Hayflick[1]

[1]*Department of Molecular and Medical Genetics, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201, USA*
[2]*Department of Neurology, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201, USA*
[3]*Department of Neurology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA*
[4]*Department of Child Neurology, Istituto Nazionale Neurologico 'Carlo Besta,' Via Celoria, 11-20133 Milan, Italy*
[5]*Neuroservices Unit, Auckland Hospital, Park Road, Auckland 1, New Zealand*
[6]*Department of Neuropediatrics, Hospital San Juan de Dios, Carretera de Esplugas s/n, 08034 Barcelona, Spain*
[7]*Department of Pediatrics, Kyoto University School of Medicine, 54 Shogoin Sakyo, Kyoto, 606 Japan*
[8]*Department of Pediatrics, Adelaide Children's Hospital, 72 King William Road, North Adelaide, South Australia 5006*
[9]*Neurophysiology Unit, Bambino Gesù Children's Hospital, Piazza Sant'onofizo 4, 00165 Rome, Italy*

Correspondence should be addressed to S.J.H.

Hallervorden-Spatz syndrome (HSS) (OMIM #234200) is a rare, autosomal recessive neurodegenerative disorder with brain iron accumulation as a prominent finding. Clinical features include extrapyramidal dysfunction, onset in childhood, and a relentlessly progressive course[1]. Histologic study reveals massive iron deposits in the basal ganglia. Systemic and cerebrospinal fluid iron levels are normal, as are plasma levels of ferritin, transferrin and ceruloplasmin. Conversely, in disorders of systemic iron overload, such as haemochromatosis, brain iron is not increased, which suggests that fundamental differences exist between brain and systemic iron metabolism and transport. In normal brain, nonhaem iron accumulates regionally and is highest in basal ganglia[2]. Pathologic brain iron accumulation is seen in common disorders, including Parkinson's disease[3,4], Alzheimer's disease[5,6] and Huntington disease[7]. In order to gain insight into normal and abnormal brain iron transport, metabolism and function, our approach was to map the gene for HSS. A primary genome scan was performed using samples from a large, consanguineous family (HS1) (see Fig. 1). While this family was immensely powerful for mapping, the region demonstrating homozygosity in all affected members spans only 4 cM, requiring very close markers in order to detect linkage. The *HSS* gene maps to an interval flanked by *D20S906* and *D20S116* on chromosome 20p12.3–p13. Linkage was confirmed in nine additional families of diverse ethnic backgrounds.

To elucidate the defect in HSS, we took a genetic approach using homozygosity mapping[8]. Others have used this method to map genes including those for Bardet-Biedl syndrome[9], familial Mediterranean fever[10], Friedreich's ataxia phenotype with selective vitamin E deficiency[11], and Hirschsprung disease[12]. We initially looked for linkage with polymorphic markers near genes encoding proteins of iron metabolism and transport, including transferrin, transferrin receptor, ferritin, and ceruloplasmin, and those encoding proteins involved in gamma amino butyric acid (GABA) metabolism and transport (glutamic acid decarboxylase, GABA transaminase and GABA receptors), based on evidence of a relation between GABA and iron metabolism[13]. As linkage to these candidate genes was excluded, we undertook a genome-wide search.

Twenty-six members from family HS1 were genotyped for 412 highly polymorphic microsatellite markers distributed along the autosomes. This highly inbred family descends from a common ancestor six generations back. Based on the average frequency of meiotic recombination (1/100 cM) and the number of meioses from founder to affecteds (6 × 2), we estimated that we would need to be within approximately 8 cM of the *HSS* gene in order to detect linkage. Prior to development of the FASTLINK programs[14], it was not possible to retain the inbreeding loops in HS1 and run even two-point analyses. For this reason, haplotypes were constructed for all markers and were analysed as an initial screening method. With markers spaced 16 cM apart, we found no evidence for linkage based on haplotype analysis, hence we screened the genome at approximately 8 cM intervals.

After excluding 95% of the autosomal genome, we established linkage to markers on chromosome 20p12.3–p13 in the HS1 family. Data from nine additional families support linkage of the HSS gene to this region with a pair-wise lod score of $Z_{max} = 13.75$ at $\theta_{max} = 0$ for the marker AFMa049yd1 (Table 1). Homogeneity analysis gave no evidence for locus heterogeneity. Homozygosity in HS1 and recombinant haplotypes in HS10, HS15, and HS16 place the *HSS* locus in a 4 cM interval between the distal and proximal markers, *D20S906* and *D20S116*, respectively (Fig. 2). Additional polymorphic markers in the interval include *D20S113*, GAAT4E12, *D20S198*, *D20S842*, *D20S181*, *D20S193*, *D20S473*, and AFM074wa9. By comparing haplotypes, we found no evidence for linkage disequilibrium. Physical mapping localizes *D20S113* to chromosome 20p13 (ref. 15) and *D20S97*, which lies within 1 cM of *D20S116*, to 20p12.3 (ref. 16). Genes that are both functional and positional candidates encode prodynorphin[17], adrenergic alpha 1D receptor[15] and phospholipase-Cβ4 (ref. 18). Other genes in the region encode FK506 binding protein, protein-tyrosine phosphatase, and sialoadhesin.

Five overlapping YACs cover the HSS critical region between *D20S906* and *D20S116*. Cloned genes in the region will be analysed for inclusion on this YAC contig and positional candidate genes and expressed sequences

Table 1 Combined pairwise lod scores for Hallervorden-Spatz syndrome and chromosome 20 markers

	Lod scores (Z) at recombination fraction (θ)								
Marker	0.00	0.01	0.05	0.1	0.2	0.3	0.4	θmax	Zmax
AFMa057vb1	−∞	3.17	3.42	3.16	2.36	1.49	0.66	0.04	3.44
D20S906	−∞	8.74	8.78	7.91	5.67	3.38	1.37	0.03	8.96
D20S113	5.70	5.57	5.06	4.39	3.03	1.72	0.60	0.00	5.70
GAAT4E12	6.20	6.06	5.49	4.75	3.29	1.97	0.88	0.00	6.20
D20S198	9.59	9.36	8.40	7.21	4.91	2.86	1.15	0.00	9.59
D20S842	13.59	13.29	12.07	10.51	7.38	4.37	1.75	0.00	13.59
AFMa049yd1	13.75	13.46	12.29	10.80	7.78	4.77	1.99	0.00	13.75
D20S181	7.61	7.42	6.68	5.76	3.98	2.38	1.04	0.00	7.61
D20S193	6.99	6.83	6.14	5.27	3.59	2.10	0.89	0.00	6.99
D20S473	8.35	8.14	7.29	6.21	4.14	2.35	0.91	0.00	8.35
AFM074wa9	−∞	2.03	2.29	2.06	1.38	0.73	0.23	0.04	2.30
D20S116	−∞	6.81	6.62	5.81	3.96	2.28	0.95	0.02	6.89
D20S867	−∞	4.69	4.68	4.10	2.76	1.55	0.59	0.02	4.81

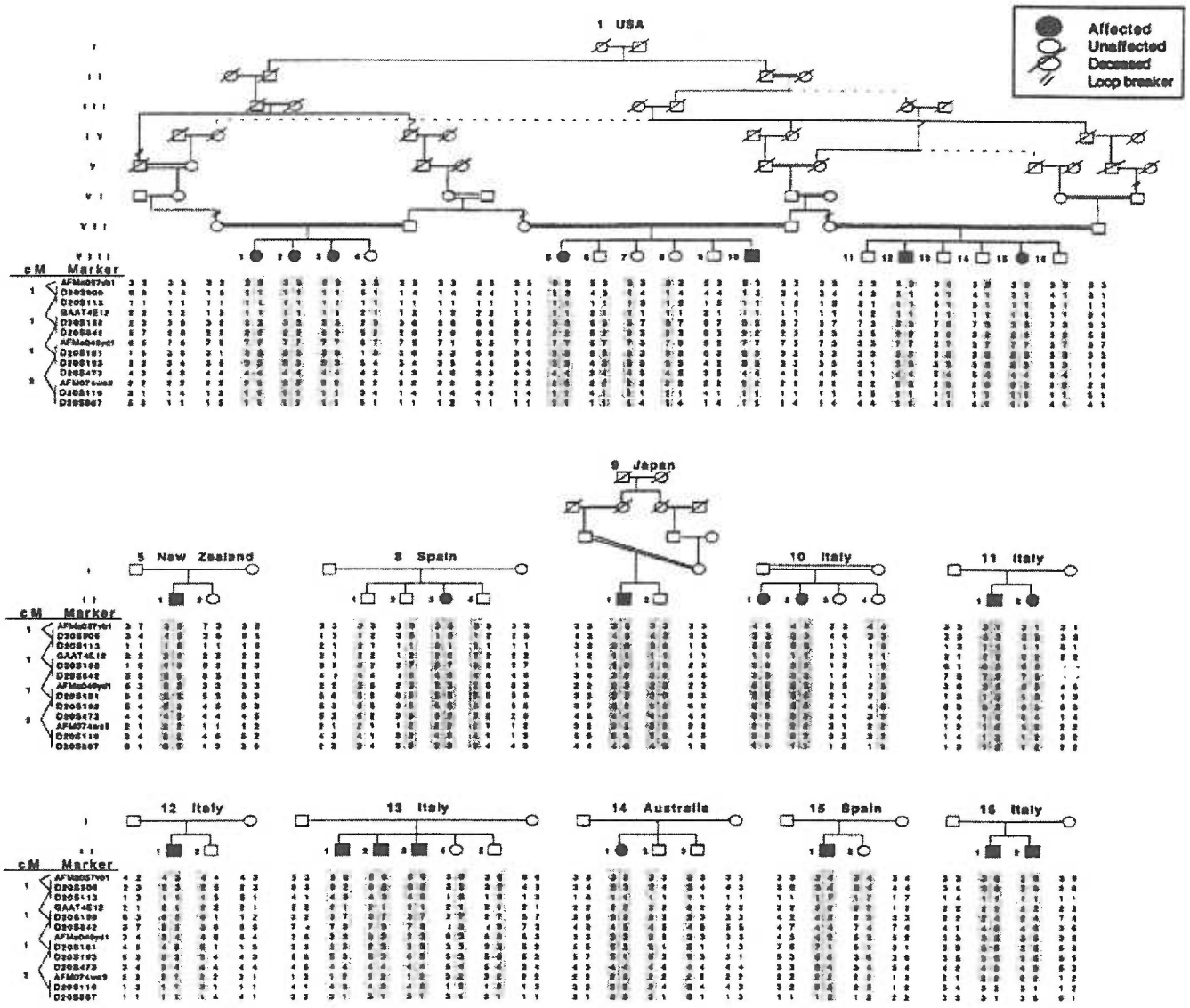


Fig. 1 HSS family pedigrees are shown with the disease-associated haplotype (shaded). Families 1 and 10 are consanguineous; the parents in family 10 are first cousins. In order to run the FASTLINK program, inbreeding loops were broken as indicated. Dashed lines indicate relationships that do not significantly impact linkage calculations and therefore were not considered.

screened for mutations. Identification of the *HSS* gene will provide the first clue to the basic defect in this neurodegenerative disorder and will contribute to our understanding of the control of normal and abnormal brain iron accumulation.

The eponym for this disorder acknowledges its delineation by Julius Hallervorden and Hugo Spatz, German neuropathologists. Shevell has brought to light Hallervorden's shameless involvement in the active euthanasia of 'mental defectives', whose brains Hallervorden collected for study[19]. We concur with appeals to rename Hallervorden-Spatz syndrome[19,20] and propose that these efforts await identification of the molecular and biochemical bases of this condition so that a rational alternative may be offered.

Methods

Clinical evaluations. Families were ascertained through clinical genetics centers and were examined by a clinical geneticist or neurologist. Detailed medical history, physical and laboratory findings and pedigree data were obtained for each family member. We have used the diagnostic criteria of Dooling *et al.*[1], further refined by Swaiman[21]. The diagnosis of Hallervorden-Spatz disease was assigned only if patients had all of the obligatory features, two or more of the corroborative features and none of the exclusionary features[21]. Obligatory features include onset during the first two decades of life, progression of signs and symptoms, and evidence of extrapyramidal dysfunction, including one or more of the following: dystonia, rigidity, choreoathetosis. Corroborative features include corticospinal tract involvement, retinitis pigmentosa and/or optic atrophy, seizures, positive family history consistent with autosomal recessive inheritance, hypodense areas involving the basal ganglia on magnetic resonance imaging, and abnormal cytosomes in circulating lymphocytes and/or sea-blue histiocytes in bone marrow. Exclusionary features include abnormalities of ceruloplasmin or copper metabolism, overt neuronal ceroid-lipofuscinosis, a family history of Huntington chorea or other dominantly inherited movement disorders, caudate atrophy, deficiency of β-hexosaminidase A or GM_1-galactosidase, nonprogressive course, and absence of extrapyramidal signs.

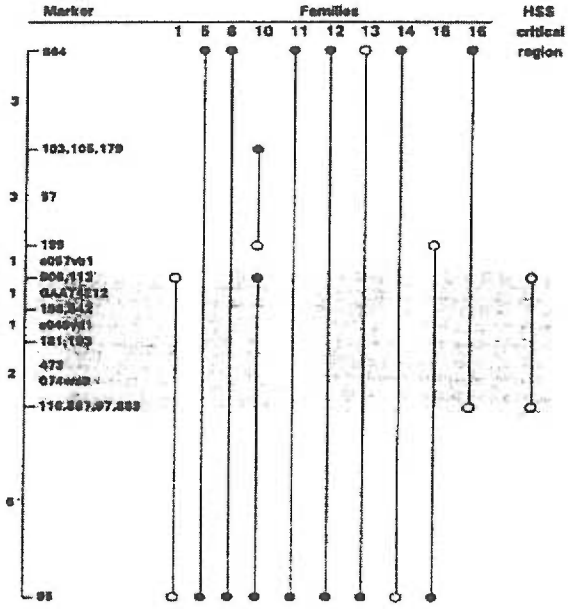


Fig. 2 The HSS critical region is defined by recombination events in the families. Vertical lines indicate the region in each family that segregates with the disease. Open circles indicate recombination with the adjacent marker(s). Filled circles indicate no recombination. Only the region of homozygosity is considered to be segregating with the gene in the HS1 family.

Ten families were included in this analysis, demonstrating wide ethnic and geographic diversity. Families were from the US, Spain, Italy, New Zealand, and Australia. The HS1 family is Amish, and their founding family member came to the US from Switzerland in 1737. In this family, each parent of an affected child is at least second-cousin to their spouse, and all are related through a common ancestor six generations back.

DNA analysis. All markers typed were PCR-based short tandem repeat, microsatellite markers (Research Genetics)[22,23]. Once candidate regions were excluded, evenly spaced markers from across the remaining regions of the autosomal genome were typed. Analysis of microsatellite markers was performed as described[24].

Linkage analysis. Linkage analysis was performed using the LINKAGE 5.1/FASTLINK 3.0P packages[16,25–28]. Autosomal recessive inheritance with complete penetrance was assumed using a population gene frequency of 10^{-3}. For lod score calculations, marker allele frequencies from the CEPH database were used. Lod scores remained significant over a range of allele frequency estimates. Linkage analysis was performed independently on each family because of the possibility of locus heterogeneity. Genetic distances between markers were based on published information[29–31], and marker order was based on published data[29,32–35] and on our genotype data. Haplotypes were constructed for all markers and were analysed for evidence of linkage.

With development of the FASTLINK programs, it became possible to retain all six inbreeding loops in the HS1 family and run two-point analyses. Based on simulations that vary the number of loops and order of loops broken, it was necessary to include only those loops from the founder to the parents of the affected children. Multipoint analysis was not feasible because of the large number of inbreeding loops in this family.

Acknowledgements

We thank the family members for their participation in this study; L. Robinson for providing key clinical information; and R. Hayflick for expert genealogical assistance. This work was supported by OHSU Foundation grant MRF9319 and by NIH grant 1K11 EY00348-01 to S.J.H.

Received 21 June; accepted 21 August 1996.

1. Dooling, E.C., Schoene, W.C. & Richardson, E.P., Jr. Hallervorden-Spatz syndrome. *Arch. Neurol.* 30, 70–83 (1974).
2. Hill, J.M. & Switzer, R.C. The regional distribution and cellular localization of iron in the rat brain. *Neuroscience* 11, 595–603 (1984).
3. Sofic, E. *et al.* Increased iron III and total iron content in postmortem substantia nigra in parkinsonian brain. *J. Neural Transmission* 74, 199–205 (1988).
4. Riederer, P. *et al.* Transition metals, ferritin, glutathione and ascorbic acid in parkinsonian brains. *J. Neurochem.* 52, 515–520 (1989).
5. Connor, J.R., Snyder, B.S., Beard, J.L., Fine, R.E. & Mufson, E.J. Regional distribution of iron and iron regulatory proteins in the brain in aging and Alzheimer's disease. *J. Neurosci. Res.* 31, 327–335 (1992).
6. Dedman, D.T. *et al.* Iron and aluminum in relation to brain ferritin in normal individuals and Alzheimer's disease and chronic renal-dialysis patients. *Biochem. J.* 287, 509–514 (1992).
7. Dexter, D.T. *et al.* Alterations in the levels of iron, ferritin and other trace metals in Parkinson's disease and other neurodegenerative diseases affecting basal ganglia. *Brain* 114, 1953–1975 (1991).
8. Lander, E.S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1567–1570 (1987).
9. Kwitek-Black, A.E. *et al.* Linkage of Bardet-Biedl syndrome to chromosome 16q and evidence for non-allelic genetic heterogeneity. *Nature Genet.* 5, 392–396 (1993).
10. Aksentijevich, I. *et al.* Refined mapping of the gene causing familial Mediterranean fever, by linkage and homozygosity studies. *Am. J. Hum. Genet.* 53, 451–61 (1993).
11. Ben Hamida, C. *et al.* Localization of Friedreich ataxia phenotype with selective vitamin E deficiency to chromosome 8q by homozygosity mapping. *Nature Genet.* 5, 195–200 (1993).
12. Puffenberger, E.G. *et al.* Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. *Hum. Mol. Genet.* 3, 1217–1225 (1994).
13. Hill, J.M. Iron concentration reduced in ventral pallidum, globus pallidus, and substantia nigra by GABA-transaminase inhibitor, gamma-vinyl GABA. *Brain Res.* 342, 18–25 (1985).
14. Schaffer, A.A. Faster linkage analysis computations for pedigrees with loops or unused alleles. *Hum. Hered.* 46 (4), 226–235 (1996).
15. Yang-Feng, T.L., Han, H., Lomasney, J.W. & Caron, M.G. Localization of the cDNA for an alpha1-adrenergic receptor subtype (ADRA1D) to chromosome band 20p13. *Cytogenet. Cell Genet.* 66, 170–171 (1994).
16. Wunderle, V. *et al.* The EUROGEM map of human chromosome 20. *Euro. J. Hum. Genet.* 2, 242–243 (1994).
17. Litt, M. *et al.* Chromosomal localization of the human proenkephalin and prodynorphin genes. *Am. J. Hum. Genet.* 42, 327–334 (1988).
18. Alvarez, R.A. *et al.* cDNA sequence and gene locus of the human retinal phosphoinositide-specific phospholipase-Cβ4 (PLCB4). *Genomics* 29, 53–61 (1995).
19. Shevell, M. Racial hygiene, active euthanasia, and Julius Hallervorden. *Neurology* 42, 2214–2219 (1992).
20. Gordon, J. Julius Hallervorden. *Neurology* 43, 1452 (1993).
21. Swaiman, K.F. Hallervorden-Spatz syndrome and brain iron metabolism. *Arch. Neurol.* 48, 1285–1293 (1991).
22. Litt, M. & Luty, J. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44, 397–401 (1989).
23. Edwards, A., Civitello, A., Hammond, H. & Caskey, C. cDNA typing and genetic mapping with trimeric and tetrameric repeats. *Am. J. Hum. Genet.* 49, 746–756 (1991).
24. Litt, M., Hauge, X. & Sharma, V. Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. *Biotechniques* 15, 280–284 (1993).
25. Lathrop, G., Lalouel, J., Julier, C. & Ott, J. Strategies for multilocus linkage analysis in humans. *Proc. Nat. Acad. Sci. USA* 381, 3443–3446 (1984).
26. Cottingham, J., R.W., Idury, R.M. & Schaffer, A.A. Faster sequential genetic linkage computations. *Am. J. Hum. Genet.* 53, 252–263 (1993).
27. Schaffer, A.A., Gupta, S.K., Shriram, K. & Cottingham, R.W. Jr. Avoiding recomputation in linkage analysis. *Hum. Genet.* 44, 225–237 (1994).
28. Terwilliger, J.D. & Ott, J. *Handbook of Human Genetic Linkage* (Johns Hopkins University Press, Baltimore, 1994).
29. Weissenbach, J. *et al.* A second generation linkage map of the human genome. *Nature* 359, 794–801 (1993).
30. Gyapay, G. *et al.* The 1993–94 Généthon human genetic linkage map. *Nature Genet.* 7, 246–339 (1994).
31. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380, 152–154 (1996).
32. Chumakov, I.M. *et al.* A YAC contig map of the human genome. *Nature* 377 Supp., 174–297 (1995).
33. Sheffield, V.C. *et al.* A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. *Hum. Mol. Genet.* 4, 1837–1844 (1995).
34. Hudson, T.J. *et al.* An STS-based map of the human genome. *Science* 270, 1945–1954 (1995).
35. Gyapay, G. *et al.* A radiation hybrid map of the human genome. *Hum. Mol. Genet.* 5, 339–346 (1996).

Homozygosity mapping of the gene for Hallervoden-Spatz syndrome to chromosome 20p12.3-p13

Todd Taylor, Michael Litt, Patricia Kramer, Massimo Pandolfo, Lucia Angelini, Nardo Nardocci, M.Pineda, Suzanne Davies, Haruo Hattori, Peter J. Flett, M. Roberta Cilio, Enrico Bertini & Susan J. Hayflick

Nature Genetics 14, 479–481 (1996).

In Fig. 1 the Hs9 pedigree was inadvertently included. The other pedigrees, the figure legened and the text are correct. Family HS9 provides preliminary evidence for locus heterogeneity in Hallervorden-Spatz syndrome. These completed data will be presented in a forthcoming manuscript.

CHAPTER 3

A GENETIC AND PHYSICAL MAP OF THE HALLERVORDEN-SPATZ SYNDROME GENE REGION ON CHROMOSOME 20p13-p12.3

A GENETIC AND PHYSICAL MAP OF CHROMOSOME 20p13-p12.3

Todd D. Taylor,* Amy Malone,* Bing Zhou,† Barbara Levinson,† Mike Litt,* Jane Gitschier,† and Susan J. Hayflick,*

*Department of Molecular and Medical Genetics, Oregon Health Sciences University, Portland, Oregon, U.S.A. †Department of Medicine, HHMI, University of California San Francisco, San Francisco, California, U.S.A.

Running title: Physical Map of Hallervorden-Spatz Syndrome Region

Reprint requests to: Dr. Susan J. Hayflick, Molecular and Medical Genetics, mail code L103, 3181 SW Sam Jackson Park Rd., Portland, OR 97201-3098. Telephone (503) 494-6866. Fax (503) 494-6886. E-mail address hayflick@ohsu.edu.

Abbreviations: HSS, Hallervorden-Spatz syndrome; PCR, polymerase chain reaction; cM, centimorgans; bp, base pair; BAC, bacterial artificial chromosome; YAC, yeast artificial chromosome; RH, radiation hybrid; EST, expressed sequence tag; STS, sequence-tagged site; NBIA1, Neurodegeneration with Brain Iron Accumulation, Type 1.

ABSTRACT

The Hallervorden-Spatz syndrome (HSS) gene was recently mapped to a 4-cM region of chromosome 20p13-p12.3, between markers D20S906 and D20S116. Currently available genetic, radiation hybrid (RH) and yeast artificial chromosome (YAC) maps from this region correlate poorly. Few markers are common to all the maps, and none of the maps is sufficiently marker dense. A high resolution radiation hybrid panel was first used to integrate markers in this region from several of maps. YACs spanning the region were typed with the same markers. A human bacterial artificial chromosome (BAC) DNA library was then screened and a preliminary contig across the HSS critical region was constructed, spanning more than 3 megabases (Mb). Given the refined mapping information, several genes and uncharacterized expressed sequence tags (ESTs) previously mapped to this region have been excluded from the HSS candidate region. This study resolves the Hallervorden-Spatz syndrome critical region to two overlapping BACs with a total size of 400 kilobases. In addition, construction of a preliminary BAC contig provides a useful tool for sequencing projects and other mapping efforts in this region.

Keywords: hallervorden-spatz syndrome/physical mapping/bacterial artificial chromosomes/yeast artificial chromosomes/microsatellite markers.

INTRODUCTION

Hallervorden-Spatz syndrome (HSS, OMIM #234200) is a rare, autosomal recessive disorder of brain iron accumulation. It is characterized by rigidity, dystonia, pigmentary retinopathy, and basal ganglia densities on MRI, specifically in the globus pallidus and pars reticulata of substantia nigra. Onset is in childhood with a progressive course leading to death by early adulthood. Using samples from a large, consanguineous, multi-generation Amish family, we were able to demonstrate linkage between the HSS gene and markers on chromosome 20p13-p12.3 (Taylor et al., 1996; Taylor et al., 1997). Homozygosity in the Amish family and obligate recombinants in three other families suggested that the HSS gene was located in a 4-cM interval between markers D20S906 and D20S116. This locus has been designated Neurodegeneration with Brain Iron Accumulation, Type 1 (NBIA1). Our next goals were to further delineate the NBIA1 critical region, by obtaining additional families and markers, and to develop a physical map of the region. Unpublished analyses (Chapter 1) have further narrowed the NBIA1 critical region to that between markers D20S473 and D20S867, which are less than 2 cM apart (CEPH, Weissenbach et al., 1992; Gyapay et al., 1994; Dib et al., 1996). This study describes the physical mapping of the Hallervorden-Spatz syndrome gene region.

Since a high-resolution map of chromosome 20p13-p12.3 was not available, we had to construct one. Publicly available maps included: genetic linkage maps (Weissenbach et al., 1992; Gyapay et al., 1994; Murray et al., 1994; Utah Marker Development Group 1995; Dib et al., 1996), radiation hybrid maps (Cox et al., 1990; Hudson et al., 1995; Schuler et al., 1996), and maps based on overlapping YACs (Cohen et al., 1993; Chumakov et al., 1995; Hudson et al., 1995). The correlation between these maps is poor

in the NBIA1 critical region. There are many inconsistencies in marker order, and not all available markers are on every map. Each of these maps has its drawbacks. Genetic maps are based on the statistical analysis of the number of meiotic recombination events between polymorphic markers, and are dependent on reliable genotyping data. Radiation hybrid maps are based on the analysis of marker-locus retention or loss in a panel of RH clones, are dependent on ease or difficulty of marker typing, and rely on even retention of donor DNA into the host cells. Because genetic and radiation hybrid mapping is statistical, the data do not suggest a unique order but rather a most likely order. Several new mapping methods are now available that alleviate most of these problems. In 1990, Green and Olson proposed a strategy whereby large segments of human DNA could be mapped using YACs (Burke et al., 1987) as the source of cloned DNA and sequence-tagged sites (STSs) (Olson et al., 1989) as the landmarks in which the map surveys. This general approach, called YAC-based STS-content mapping (Green and Olson, 1990a; Green and Green, 1991), has been used to map numerous targeted regions of the human genome. YAC clones are often chimeric, and are often internally deleted or rearranged, so BACs and P1 artificial chromosomes (PACs) are now being substituted as the sources for cloned DNA using the same approach. BAC and PAC (Shizuya et al., 1992; Ioannou et al., 1994) clones have low frequencies of chimerism and are more stable than YACs or cosmids. BAC and P1 maps are not statistically based, can be used to integrate all types of markers, and require simple plus or minus screening by PCR or hybridization.

We now describe the construction of a high-resolution, BAC- and YAC-based STS-content map across approximately 3 megabases of human chromosome 20p13-p12.3. Results of the radiation hybrid mapping analysis are detailed in Chapter 1.

Common Name	Source	Type	Aliases	Length	Accession	Forward Primer	Reverse Primer
110K24-60	OHSU	STS	AM060	115	NA	ATACACAGACACCAATGCAAGG	GTGGATCTGCATTTCTTTAGGG
142L14-75	OHSU	STS	AM075	289	NA	GCTGGGTTCCTTTGCATAAA	GCAGTTTGTTTGAAGTTCAAGG
256O7-63	OHSU	STS	AM063	260	NA	CTCTGAAATAGTGCCAGGTACA	TGGAATCCAATATTAGGCATCC
271B20-69	OHSU	STS	AM069	200	NA	CCTCCACTCTCCTGCTATGC	AATGGAGGGCTCAGTGTCC
271B20-70	OHSU	STS	AM070	251	NA	TGATTTTAAAGATGGGCAACG	GACAGAAACCACTCATCTTCCC
312F20-65	OHSU	STS	AM065	164	NA	TGTCCTTTTTGCAAGCAGTG	TCGTGAGATGTCCCTTTTCC
355G8-1	UCSF	STS	BZ01	400	NA	CCTGAAACAGAAGCCCTGCTAACC	TGGAAGTAGGTTAGCATAAAAC
355G8-2	UCSF	STS	BZ02	400	NA	ACTTATTCTGTAAGCTTTCTCTAT	CAGTCATATAAAAGTATAGCACAT
39L21-71	OHSU	STS	AM071	155	NA	CACGTTAACATGATGGTTTTGC	AGCTCCTTTGCTTTCACCAC
39L21-72	OHSU	STS	AM072	181	NA	AGGCCTCTGCTAATATCAGGG	CCTCAGCTGATGAGAAAAACG
AFM074wa9	Genethon	VNTR	w1699, stSG-20213	163	Z66604	TTTGAGAACTCTATACGGGAGTCT	CAGCCACCTCAGTGTGTTTT
AFMa049yd1	Genethon	VNTR	w857	269	Z67154	CACCAAGCACTGACAATC	CCTGGCCCCTGTTAAT
CENPB	GenBank	EST		153	X55039	ATCTGTCCACTCCCCTTTTG	ACCCTCTCAACACTTGCCTC
D20S97	Genethon	VNTR	AFM036ya3	272	Z16462	GGGAGATGGAGAGTTGTTGC	TATACGCACCCATACCACCA
D20S116	Genethon	VNTR	AFM248td1, stSG-10719	111	Z17107	TGACCACAGGGGTTAATG	CAGGACTCAGTGCACCAG
D20S181	Genethon	VNTR	AFM240zf4	161	Z23780	ATCCCTCTAAGCATGGGC	GGGTCTCTGTCAATGGGT
D20S193	Genethon	VNTR	AFM308we1	152	Z24264	GCATCCCTGGGCTAACTG	GGAATCTTTTGGTGCGTG
D20S198	Genethon	VNTR	AFM333xe5	234	Z24439	AGTGAGCCCAGGTCGT	TAGGAACATCATGTAACCCAGA
D20S437	Utah	VNTR	UT235	415	L29932	GCAGTGAGTTCATATGGCTA	GTTTCTCTGTTCAGCACTTC
D20S473	CHLC	VNTR	ATA21E04	180	G08041	TCATGAGCTAAATATTACTCAGTGC	CTTATAGCTTTTTTCAAATGATCTG
D20S482	CHLC	VNTR	GATA51D03, HS052191	151	G08052	AGCCTCCATAACCACATGAA	GAACCTAAAACTCTAAGGAAGCG
D20S596E	WICGR	EST	stSG3058	123	Z43546	GCCTCAGCCAGTTTTGCC	GTCCCAACTCCTCTCCTC
D20S619	WICGR	STS	WI-5517, MR10442	177	G03667	GGGTTGTTTACCTGGCAAGA	TGGAGTATGGACAGGGATGG
D20S731	WICGR	EST	PTPRA, WI-8798, EST108539	117	T55256, G06125	TTGGTGCCAGAGCTACCTG	TTTGGATTGATATCGTGAAATCC
D20S751	WICGR	STS	WI-4689, MR8061	124	G04693	CACTTGAAGTATTATTCATGCAGTC	AACATGGGCAAATTCTTCATG
D20S752	WICGR	STS	WI-4876, MR9678, stSG-10877	157	G04713	TGATGTTAGTTAAATGTTCCACAAG	TCCTGGCATATGTTCAATTCC
D20S828	SHGC	STS	SHGC-2800, stSG-20468	167	G07505	TCTTAGGAGCAGAAGCAGCATAG	AACCACCCACAAAAACCACCTTAC
D20S842	Genethon	VNTR	AFMa175vb1	163	Z52353	AGCGCACAGCCTTCAT	AGCTTCCANCCATTCAT
D20S849	Genethon	VNTR	AFMa217zh9, w5723	224	Z52536	GCTGCTCATACCTTCAGGA	TGGCACACGTTTACCTACA
D20S867	Genethon	VNTR	AFMb026xh5	277	Z53139	TACAGGCATGAGCCAT	TCTGAAATAACTTGTTTTGTGA
D20S889	Genethon	VNTR	AFM234tf10, w2437	279	Z51190	GGTTTGGTGGAATCCTCTC	CATCTTTCAAATGGGATAATGG
D20S895	Genethon	VNTR	AFMb352xd9, w6846	221	Z53825	CCCAGGGAGGTAAAGGTT	GTCAGGCTACATCAGCAAA
D20S1012	WICGR	EST	CDC25B, WI-7686, UTR-03283	336	G06651, S78187	CCTATTTCAGTGTTACCTGTGTGC	GTTTGGGTATGCAAGGCACT
D20S1032	WICGR	EST	SNRPB, WI-9238, UTR-03786	109	G07193, M34081	GGAAGTAGCTCCGCAGAGG	GGAAACCAGACAATCCCATG
D20S1116	SHGC	STS	SHGC-14786	150	G14942	AGCTCTAACCTGCCTTGGAT	TCTGAAGGGCATTTAGGGG
OXT	GenBank	EST	HUMOTNPI	136	M11186	CGCTGAAACTTGATGGCTC	CCCATTTCTGGCACTTAGAC
PCNA	GenBank	EST		221	J04718	AAAAGCCACTCCACTCTCTTC	GATCCTTCTTCATCCTCGATC
SGC31723	SHGC	EST	EST29426, RH23553	150	G29331	TTTAATTTCCTGTGACCCTTTTA	GTTTATACCATCCAACTGTGAACTC
SGC32955	WICGR	EST	EST82406	127	T32373	TTACACGTTTGAGTCAATTACATGC	TTTTAAACATTCGCAGTTTTTTG
SHGC-2666	SHGC	EST	FLT	218	D11843	ATCTCTGTGAACTTCCTGGAGAC	TTAGCCAGAAGCCCTATTACTTTG
SHGC-30809	SHGC	EST		150	G27198	CTCAGCATTGCTCTAGGGGA	CTGACAAGATTCAACGGGCC
SN	GenBank	EST		600	NA	CTCTGTGTTCCAGCCCTGCA	CTCGCCCATGCTCTGCTTAC
stSG-408	WICGR	EST	EST50864	112	D20236	TCATGCACCATGACCCAC	CGCTGGAGCAGAATGAAACT
stSG-2589	WICGR	EST	EST66078	147	Z39952	CAGCATGCATCATCCAAAGT	CTTTTTGGACTGAAGTGCAGG
stSG-4244	WICGR	EST		127	H49371	TGACCCTGAAGGCAGAGACT	CTAGGCAGGAAGACCCCTG
stSG-8000	WICGR	EST		128	T65776	CAGCAGCTACACTGGAACCA	TCCTATGCACCAGTAAGGCC
stSG-9697	WICGR	EST	A007003	162	F13536	GATTAATAAGCAGTGAGTTACTG	CTAAAGCTAGAGCTTTCATTTA
stSG-9792	WICGR	EST	A007Q34	155	H00291	ACTCTTAATGATACTCAGATGAC	CACTTGGTCAGAAACTTC
stSG-10203	WICGR	EST	A008E19	101	R49090	CTTTGTGGTTTTAGTGAAGAAGT	TACTCTTAAGAGGAGTGTGTGTG
TIGR-A005O05	TIGR	EST	THC95046	239	F08919	TCAGAAGGAAGTTTGTATTATAG	TTAGCATCTATTACACTCATGTA
TIGR-A008V36	TIGR	EST	THC89272	221	T80373	TCTTCTTTACAGTGGTAGGTTTT	AGGACTGAATCCAAATCC
WI-4715	WICGR	STS	MR8569	200	NA	CTCAATGGCTTCCATGTCCT	CTCGGCAAGAGTGTGCCT
WI-9015	WICGR	EST	PTPRA, UTR-03102	123	G07079, M34668	TTTGGATTGATATCGTGAAATCC	TTCTTAGTTGGTGCCGAGCT
WI-11482	WICGR	EST	EST159813	127	T89692	TTGTTAAAATGCAAACCAATTCA	ACATGGTGGAAAATGAATGTTG
WI-11626	WICGR	EST	EST179747	150	R09451	TTGATTTTACTAAGGTCTTCCACTG	TCATTGCCAATTTGTTCTTG
WI-13669	WICGR	EST	EST230884	130	R44479	CGTTTTATTTCTTTGGGAAAGG	AAAATATGAACAGACTCAGCTTTGG
WI-16193	WICGR	EST	EST285473	128	H24962	ACCTTACATGGGTATGGACAGG	ATGATGATGATTCTCCCCCA
WI-16594	WICGR	EST	EST91725	125	T40921	TGTAGTCCCAACCTCCATCC	TTACCTAAGAGTGGGACACAACC
WI-17847	WICGR	EST	EST327172, stSG-20158	133	R88736	TGCTTTAGTTCATAGTTGAAGTCCA	ACTTAATGGCAATTCGGTGC
WI-18557	WICGR	EST	EST91360	128	T40556	CACTCCTCTGGAGTACTACTGCTG	GAAGCGGCTGCTGAGATTAC
WI-30160	WICGR	EST	EST351648, SGC30160	125	H50438	TTGAACCAAGGCCCTAACAG	CAAGAGGAAATGTCACAGAAGC

Table 7. Marker summary. NA - not available yet.

MATERIALS AND METHODS

Chromosome 20 STSs. Sixty-one PCR-typable chromosome 20 STSs were used (**Table 7**) to screen various libraries. Markers were obtained from either Research Genetics, Inc., or Life Technologies (Rockville MD, USA). Fifteen short tandem repeat microsatellite markers came from several groups: Généthon (Weissenbach et al., 1992; Gyapay et al., 1994; Dib et al., 1996); the Cooperative Human Linkage Center (CHLC) (Murray et al., 1994); and Utah (Utah Marker Development Group, 1995). Ten STSs were generated from BAC end sequences in this study. Four primer sets were developed from GenBank gene sequences. The remaining 26 ESTs and six STSs came from several groups: Stanford Human Genome Center (SHGC), Whitehead Institute/MIT Center for Genome Research (WICGR), and The Institute for Genomic Research (TIGR) (Hudson et al., 1995; Schuler et al., 1996).

PCR. PCR amplification for most markers was performed in a total volume of 14 µl with 20-40 ng template DNA, 3.5 pmol each primer, 1.5 mM MgCl2, 2.8 mM dNTPs, 50 mM KCl, 10 mM Tris-Cl, pH 8.3, 0.25 mM spermidine and 0.35 U of Ampli*Taq* DNA Polymerase (Perkin-Elmer Cetus Corp.). In a typical experiment, three-temperature "touchdown" PCR amplification (Don et al., 1991) was performed in a Gene Amp PCR System 9600 thermocycler (Perkin-Elmer Cetus Corp.). An initial denaturation step of 4 minutes at 94°C was used, followed by an annealing temperature of 70°C which was decreased by one degree in each of the first fifteen cycles, then maintained at 55°C for 20-25 more cycles. Each cycle consisted of a 15-second 94°C denaturing step, a 30-second annealing step and a 30-second 72°C extension step. This was followed by a final 5-minute extension step at 72°C, after which the temperature was reduced to 15°C until

the reaction tubes were removed from the apparatus.

Genomic libraries and screening. The human BAC plasmid DNA pools distributed by Research Genetics, Inc. (Release II, Shizuya et al., 1992), were screened by PCR. These pools represent a 4-5X coverage of the whole human genome. The configuration of the pools (48 superpools that each correspond to 8 plate, 16 row, and 24 column pools) results in the exact well location of a BAC clone in just 96 PCR reactions. Individual clones were also obtained from Research Genetics, Inc., in either agar stabs or 384-well microtiter plates. These clones have an average insert size of 130 kb, with a range of 90-300 kb. Clones were grown in 500 ml LB broth containing 12.5 μ g/ml chloramphenicol and DNA was extracted using the QIAGEN Maxi Plasmid Purification protocol (QIAGEN Plasmid Handbook, February 1995).

BAC clone end sequences. BAC clones were obtained from Research Genetics, Inc. Clone ends were obtained by direct sequencing 0.5 μg of BAC DNA using an ABI377 sequencing unit and dye terminator chemistry. 3.2 pmols of the following primers were used: pUC/M13 Forward(-47) 5'-d(CGCCAGGGTTTTCCCAGTCACGAC)-3' or pUC/M13 Reverse 5'-d(TCACACAGGAAACAGCTATGAC)-3'. Sequences obtained were tested using the BLAST program available through NCBI to identify repetitive elements or any other homology. Unique sequences were selected and PCR primer pairs were designed with the Primer program (Version 0.5; Whitehead Institute). Potential primer pairs were then examined for robustness and specificity with the Macintosh program Amplify (William Engels, University of Wisconsin).

YAC Clones. YAC clones from the CEPH/Whitehead Institute map (Chumakov et

al., 1995) were obtained from Research Genetics, Inc., and Genome Systems, Inc. (St. Louis MO, USA), and screened by PCR. Clones were received in agar stabs, and DNA was extracted by standard methods (Hoffman, 1993).

Radiation hybrid mapping panel. The high resolution Stanford TNG3 radiation hybrid panel distributed by Research Genetics, Inc., was screened using 10 µl PCR reactions. This panel consists of 90 radiation hybrid clones of the whole human genome, plus two control DNAs (RM donor and A23 recipient). The human lymphoblastoid cell line (donor; RM) was exposed to 50,000 rad of X-rays and fused with nonirradiated thymidine-kinase deficient hamster recipient cells (A23), creating a panel of independent somatic hybrid clones with 60-kb resolution. Software for the analysis of results is available free of charge via the University of Michigan's Statistical Genetics Software web site (http://www.sph.umich.edu/group/statgen/software). Analysis was performed with the programs RH2PT and RHMINBRK from the RHMAP software package (Boehnke et al., 1991; Boehnke, 1992).

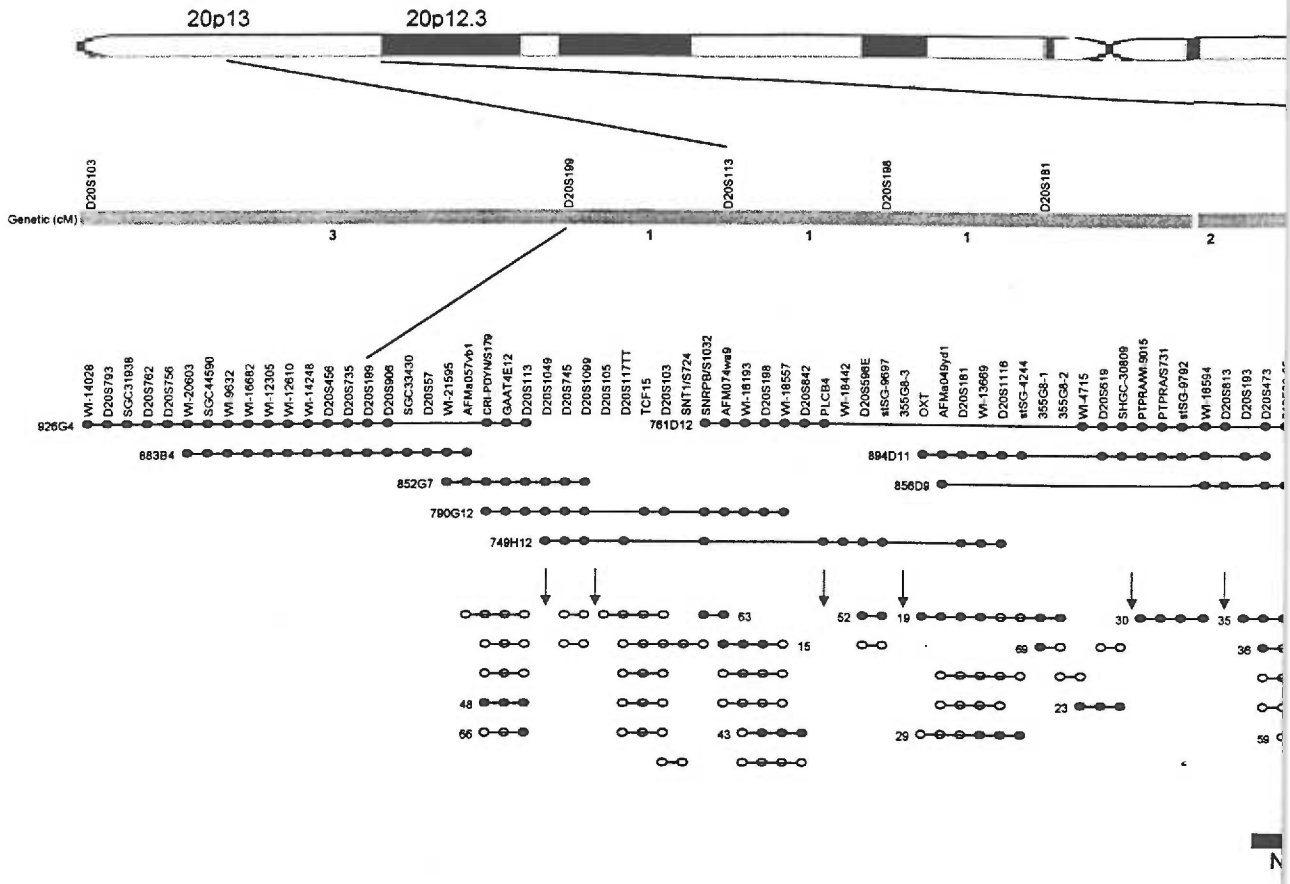


Figure 11. Physical map for portion of chromosome 20p13-p12.3. The cytogenetic map is at the top, followed by the genetic, YAC-STS content, a been confidently placed on a clone. Open circles indicate markers that are likely to be on the clone (from superpool screening) but have not yet bee name (see Table 8). The NBIA1 critical region is indicated at the bottom by the horizontal bar.

	YACs	790G12	749H12	761D12	894D11	856D9	753G9	BACs	345M/N4	39P5	24304	341C9	355G8	162I21	277N15	168H22	312F20	120C11	105L22	13N6	142L14/17	271B20	134G23	37G20	270F18	256O7	290A14	66F21	112J7	355G11	291D3	110K24	108A15	287B15	17K7	282H15	24P12	39L21
	Size(kb)	~1080		1330	~1590	790	1380	Size(kb)			69						220	180		220						74						72					
Markers								AKA	63	15	43	52	19	29	23	30	35	36	59	28	32	2	3	14	10	38	41	11	45	18	42	46	47	61	53	60	56	17
SNRPB/S1032		+	+	+					+																												
AFM074wa9		+	-	+					+	+																											
WI-16193		+	-	+						+	*																										
D20S198		+	-	+						+	+																										
WI-18557		+	-	+						*	+																										
D20S842			-	+							+																										
D20S596E			+	-								+																									
stSG-9697			+	-								+																									
OXT			-	-	+								+	*																							
AFMa049yd1			-	-	+	+							+	*																							
D20S181			+	-	+	-							+	*																							
WI-13669			+	-	+	-							+	+																							
D20S1116			+	-	+	-							*	+																							
stSG-4244				-	+	-							*	+																							
355G8-1				-	-	-							+																								
355G8-2				-	-	-							+		*																						
WI-4715				+	-	-									+																						
D20S619				+	+	-									+																						
SHGC-30809				+	+	-									*																						
PTPRA/WI-9015				+	+	-										+																					
PTPRA/S731				+	+	-										+																					
stSG-9792				+	+	-										+																					
WI-16594				+	+	+										+																					
D20S193				-	+	-											+																				
D20S473				+	+	+											+	+																			
312F20-65				+		+											+	*	*																		
D20S828				+		+												+	+	+																	
SN						+														+	+																
D20S867						+														+	+																
D20S752						+														+	+																
A005O05						+															+																
stSG-408						+															+																
142L14-75						-															*																
271B20-69						+	+															+															
D20S116						+	+															+															
SHGC-2666						+	+															+	+	*													
271B20-70						+	+															+	*	*													
D20S889						+	+																+	+	+												
256O7-63						+	+																*	*	*	+	*										
stSG-8000						+	-																	*	*	+	*	*										
SGC32955						+	+																	+	+	+	+	+										
WI-17847						+	+																			+		*										
WI-30160							-																			+			*									
CENPB							-																			+			*									
CDC25B/S1012							-																			+			+									
D20S482							+																							+	+							
D20S97							+																							+	*	+						
D20S437							+																							*	*	+						
110K24-60							+																							*	*	+						
WI-11482																																	+	+				
stSG-2589																																	+	*	+			
A008V36																																	*	*	+			
SGC31723																																		+	+			
D20S751																																			+			
D20S849																																				+		
stSG-10203																																				+	+	
D20S895																																					+	+
WI-11626																																					*	+
PCNA																																					*	+
39L21-71																																						+
39L21-72																																						+

Table 8. Summary of the chromosome 20p13-p12.3 physical map. 61 markers used in this study are listed in the first column. The top rows indicate the names and insert sizes (when determined) of 6 YACs and 30 BACs. Only the BACs (and corresponding markers and YACs) where an exact marker address was determined are shown. Plus signs indicate markers that have been confidently placed on clones. Stars indicate markers that have not been verified yet. Minus signs indicate likely deletions in the YACs.

RESULTS AND DISCUSSION

The results of screening both the YAC clones and BAC DNA pools with markers from chromosome 20p13-p12.3 are shown in **Figure 11**. The YACs were screened with 115 markers, and the BAC pools with 93 markers. There were on average 4.5 positive hits per marker screened with the BAC DNA pools. Most of the markers screened with the BAC pools could be included in eleven "mini" contigs, with ten intervening gaps. Most markers could be ordered just by the number of positive superpool hits they had in common. A second round of screening for some of the positive hits was performed to identify their exact clone addresses.

The contig was initially assembled based on BAC superpool screening information. Many markers were ordered based solely on the number of common superpool hits they shared. The more positive superpool hits two markers had in common, the more likely they had at least one BAC in common. This likelihood was estimated from the number of clones in the library and, specifically, from the number of clones in each superpool. The human BAC plasmid DNA pools (Release II, Shizuya et al., 1992) distributed by Research Genetics, Inc., represent a 4-5X coverage of the whole human genome. There are 147,456 unique clones in these pools. Each superpool (48) contains 3072 clones (8 plates x 384 clones per plate). Given two markers have the same superpool hit in common, the probability that they are not on the same clone equals 3071/3072 (0.99967):

$$P(\text{2 markers not on same clone} \mid \text{1 common superpool hit}) = P(G) = 3071/3072 = 0.99967$$

If one were to take any two markers not previously known to segregate together in the genome, the probability that these two markers are on at least one common clone,

given that they have only one common superpool hit, is:

$$P(\text{2 markers on at least one common clone} \mid \text{1 common superpool hit}) = 1 - P(G) = 3.26\ 10^{-4}$$

The probability that these two markers are on at least one common clone, given that they have 2 common superpool hits, is:

$$P(\text{2 markers on at least one common clone} \mid \text{2 common superpool hits}) = 1 - (P(G))^2 = 6.51\ 10^{-4}$$

The probability that these two markers are on at least one common clone, given that they have 5 common superpool hits, is:

$$P(\text{2 markers on at least one common clone} \mid \text{5 common superpool hits}) = 1 - (P(G))^5 = 1.63\ 10^{-3}$$

In this study, we were not considering markers that are randomly dispersed throughout the genome. We were looking at an approximately 5 Mb subset of the genome with markers that are all known, *a priori*, to be mapped to chromosome 20p. Therefore, we were only looking at a subset of clones within the superpools. The human genome is approximately $3x10^9$ bp. Assuming library coverage of the genome is 4.5X (which matches the average number of hits in this study), each BAC clone is ~91,552 kb ($3x10^9$ bp / 147,456 clones). So, in a 5 Mb region of the genome, with 4.5X coverage, there will be approximately 246 clones ((5 Mb / 91,552 kb) x 4.5 clone coverage). If we assume that these 246 clones from our region of interest are distributed evenly among the 48 superpools, then there are only about five clones per superpool (246 clones / 48 superpools). This then increases the probability that two markers are on at least one common clone, given a certain number of common superpool hits. Therefore, given two markers have the same superpool hit in common, the probability that they are not on the

same clone equals 4/5 (80%):

$$P(\text{2 markers not on same clone} \mid \text{1 common superpool hit}) = P(S) = 4/5 = 0.8$$

For two markers that are known to map to the same 5 Mb region, the probability these two markers are on at least one common clone, given they have only one common superpool hit, is:

$$P(\text{2 markers on at least one common clone} \mid \text{1 common superpool hit}) = 1 - P(S) = 0.2 = 20\%$$

and, the probability that these two markers are on at least one common clone, given that they have 2 common superpool hits, is:

$$P(\text{2 markers on at least one common clone} \mid \text{2 common superpool hits}) = 1 - (P(S))^2 = 36\%$$

and so on:

$$P(\text{2 markers on at least one common clone} \mid \text{3 common superpool hits}) = 1 - (P(S))^3 = 49\%$$

$$P(\text{2 markers on at least one common clone} \mid \text{4 common superpool hits}) = 1 - (P(S))^4 = 67\%$$

$$P(\text{2 markers on at least one common clone} \mid \text{5 common superpool hits}) = 1 - (P(S))^5 = 74\%$$

Clearly, the probability that two markers, from the same chromosomal region, are on at least one common BAC clone rises dramatically with the number of common superpool hits they share. Ordering the markers by their common BAC superpool hits led to a reduction in the number of secondary screenings that had to be performed. In many cases, only a few superpools had to be followed up to identify overlapping clones or markers that were on the same clone.

Exact BAC clone addresses were determined by subsequent screening of the plate, row, and column pools. Where there were ambiguities in marker or clone order, information from STS-content screening of YAC clones and screening of a high-resolution radiation hybrid panel was reviewed. There are still several gaps in the BAC contig that must be resolved by additional rounds of end sequencing of the gap-flanking BACs, primer design, and rescreening of the superpools. Screening the more recent version of the Research Genetics, Inc. human BAC pools (release IV) , which now claims a 12X coverage of the whole human genome, may help to close these gaps.

The physical mapping data for chromosome 20p13-p12.3 is summarized in **Table 8**. All of the markers and YACs used in this study are shown, but only those BACs where an exact clone address was identified are displayed. Most of the gaps in the BAC contig are covered by one or more YACs except on the proximal end. It is evident that some of the YACs are deleted and, therefore, are not reliable sources of coverage for gaps in the BAC contig. The YACs are also problematic in that many are chimeric and their sizes are uncertain.

We have developed a high-resolution, integrated BAC and YAC contig across a portion of chromosome 20p13-p12.3. Using all of the available mapping data for this part of the chromosome, we were able to construct a contig spanning approximately 3 megabases. Several groups, including Généthon, the Whitehead Institute, CHLC, and the Stanford Human Genome Center have generated maps of this region. However, there are significant discrepancies among these maps and few markers are on all of the maps. We were able to use these maps along with the family haplotype data to integrate most of the markers into a high-resolution map. Direct comparison of the maps is difficult, but there

is general consistency between the map we generated in this study and the maps from the other groups. Filling the remaining gaps, identifying additional BAC clones for each marker, and fingerprinting of overlapping BAC clones will help strengthen the contig and make it more useful for sequencing efforts. Fingerprinting the clones will confirm contig integrity and help define a minimal tiling path across the region. This first-generation BAC STS-content map of chromosome 20p13-p12.3 provides an essential resource for genetic and molecular studies of the region.

The 20p13-p12.3 region was of particular interest to our group because it harbors the gene for Hallervorden-Spatz syndrome (Taylor et al., 1996; Taylor et al., 1997). Since this study began, we continued to genotype additional markers and families (Chapter 1). Two families, with obligate recombinant events between the NBIA1 gene and markers from the 20p13-p12.3 region, have helped narrow the NBIA1 critical region to that between markers D20S473 and D20S867. Both markers lie on YAC 856D9 (790 kb), which is known to be chimeric. This interval is also spanned by two overlapping BACs (120C11 and 13N6; 180 kb and 220 kb, respectively) with a total size of 400 kb. Development of additional markers within this region may help narrow the Hallervorden-Spatz syndrome gene critical region.

EPILOG

This work describes the genetic and physical mapping of the gene for Hallervorden-Spatz syndrome. HSS is a rare, autosomal recessive disorder with brain iron accumulation as a prominent finding. HSS is characterized by rigidity, dystonia, pigmentary retinopathy, onset in childhood, and a progressive course leading to death by early adulthood. Little is known about the biochemical or cellular defects in this disease. Iron accumulates intra- and extracellulary in the basal ganglia, specifically in the globus pallidus and substantia nigra pars reticulata. Genes encoding proteins for GABA and iron metabolism, and the proteins they interact with, as well as loci associated with isolated RP and other forms of syndromic RP were considered candidate genes for HSS.

To elucidate the defect in HSS, we first took a genetic approach using homozygosity mapping (Lander and Botstein, 1987). To reduce the possibility of locus heterogeneity, linkage analysis to the NBIA1 gene was performed using samples from just one family. This family, HS1, is a large, inbred Amish family with multiple consanguineous unions. Linkage analysis was performed using the LINKAGE 5.1/FASTLINK program packages (Lathrop et al, 1984; Cottingham et al, 1993; Schäffer et al., 1994; Terwilliger and Ott, 1994; Schäffer, 1996). For lod score calculations (Morton, 1955), marker allele frequencies were assumed to be equal, since reliable population frequencies for many of the markers was not available. Because of the large size of the HS1 family and the multiple number of consanguineous unions, there was a strong potential to find linkage. One problem with there being so many consanguineous unions in this family, was that there were many loops in the pedigree that slowed linkage calculations. By experimenting with the number of loops in the HS1 family, we found that it was only

necessary to include three loops in the linkage analysis. This was the minimum number of loops required to connect the founding generation to each of the carrier parents in the seventh generation. No information was lost by leaving out additional loops, and computation time was dramatically reduced.

Highly polymorphic microsatellite markers (Litt and Luty, 1989; Weber and May, 1989) were typed near candidate genes encoding proteins of iron homeostasis and pigmentary retinopathy. Once those regions were excluded, we performed a random screen of the entire autosomal genome. Markers were spaced on average every 40 cM and had an average heterozygosity of about 70%. When no significant evidence for linkage was found for the first set of markers (lod score, z, greater than 3), we increased the density of the markers to one every 20 cM. Eventually, markers were run in any region where the gap was greater than 16 cM. Two-point linkage analysis was used to exclude or confirm linkage. In regions where linkage analysis alone was insufficient to exclude the area between two closely linked markers, haplotype information was used to exclude the region. When haplotype information could not be used to exclude these regions, at least one additional marker was run between the two flanking markers. Multipoint linkage analysis could not be used because of the complexity of the HS1 family.

After running about 400 markers and excluding about 90% of the autosomal genome, we found significant linkage to marker D20S199 ($z = 5.92$ at $\theta = 0.01$) on chromosome 20p13-p12.3. Analysis of nine other families supported linkage of the NBIA1 gene to this region with a total maximum two-point lod score of 13.75 at $\theta = 0$ for marker AFMa049yd1. Only one family, HS9, with classical HSS was not linked to this region.

Because the allele frequencies in the CEPH database closely coincided with the allele frequencies in our families, the CEPH allele frequencies were used to calculate the lod scores. When equal allele frequencies were used instead, the lod scores were inflated by an average of 0.59 at $\theta = 0$. This locus for Hallervorden-Spatz syndrome has been designated Neurodegeneration with Brain Iron Accumulation, Type 1 (NBIA1).

Haplotypes spanning the NBIA1 critical region were constructed for all families. Markers from several maps were genotyped and ordered based on published data (NIH/CEPH Collaborative Mapping Group, 1992; Weissenbach et al., 1992; Chumakov et al., 1995; Sheffield et al., 1995; Hudson et al., 1995; Gyapay et al., 1996). There were often discrepancies in the order of the markers between these maps. A few double-recombinants were found in our haplotype data that were resolved by either re-genotyping the individuals in question or by altering the order of the markers. From the initial set of families, the closest distal and proximal markers to the NBIA1 gene locus were defined by obligate recombinants in the HS1 and HS16 families, respectively. The NBIA1 critical region was narrowed to a 4-cM interval between markers D20S906 and D20S116.

Several previously cloned genes mapped to the NBIA1 critical region, including prodynorphin (Litt et al., 1988; Summar et al., 1990) and sialoadhesin (Mucklow, 1995; Kelm, 1996). Without better mapping information, neither of these genes could be excluded from the NBIA1 critical region. Many uncharacterized ESTs had also been mapped to the NBIA1 critical region. Prodynorphin is expressed in the basal ganglia, and encodes a neuropeptide precursor that is present in regions of high iron accumulation in HSS patients. The NIRCA method (Goldrick et al., 1996) was used to perform mutation

analysis for several HSS individuals, but with no significant findings. Sialoadhesin (OMIM #600751) is a cell-cell interaction molecule expressed by subpopulations of tissue macrophages (Mucklow et al., 1995). SN contains 17 immunoglobulin-like domains and is structurally related to three other members of the sialoadhesin family. Studies in mice implicate that the mouse homologue (Sn) may play a role in the development of myeloid cells in bone marrow and in the trafficking of leukocytes in lymphatic organs (Kelm et al., 1997). One of the other members of the sialoadhesin family, myelin-associated glycoprotein, is expressed in myelin of oligodendrocytes and Schwann cells and specifically targets neurons and oligodendrocytes. It may be that in humans, SN interacts with or modulates synaptic transmissions in a manner which when disrupted leads to a cascade of events like those in the pathophysiology of HSS. Our collaborators at UCSF are preparing to screen several of the HSS individuals by Southern and Northern blot analysis, using a cDNA probe for sialoadhesin.

Two other methods were used to try to narrow the NBIA1 critical region. Linkage disequilibrium analysis was done with programs DISLAMB (single locus analysis) and DISMULT (multilocus analysis) (Terwilliger, 1995). Neither analysis returned any significant results, most likely due to the small sample size. Linkage disequilibrium analysis requires a large number of families from the same ethnic background or genetic isolate; in contrast, we had only seven Italian families of unknown geographical distribution. Shared haplotype analysis was performed by observing where families from the same country (except the USA and Canada) shared three or more adjacent alleles in the original 4-cM NBIA1 critical interval. All of the families, except for the two Spanish families, shared at least one chromosomal haplotype with at least one other family. While

there was no common allele or haplotype that was shared within any particular group, they did share segments that either flanked or spanned the NBIA1 critical region (between markers D20S473 and D20S867). This analysis did not help narrow the NBIA1 critical region, but gave supporting evidence that the NBIA1 region was between these two markers. It also strongly suggested that several of the families might be distantly related to one another, due to a significant amount of allele sharing.

As additional markers (and families with classical HSS) became available, they were genotyped and incorporated into the analyses. The NBIA1 critical region was further narrowed by the inclusion of HS50 family. The closest distal marker became D20S473 and the closest proximal marker became D20S867, due to obligate recombinants by the HS50 and HS16 families, respectively. With improved haplotype information, the order for markers D20S867 and D20S116 reversed. The HS16 family still defined the proximal end, but the closest proximal marker became D20S867. Markers D20S473 and D20S867 were less than 2 cM apart according to the CEPH maps.

Since many of the candidate genes on chromosome 20p13-p12.3 could not be included or excluded from the NBIA1 critical region solely on the haplotype data, we began to construct a physical map for the region. No high-resolution map of chromosome 20p13-p12.3 was available that integrated all types of markers. The correlation between numerous available maps of chromosome 20 was poor (Weissenbach et al., 1992; Gyapay et al., 1994; Murray et al., 1994; Utah Marker Development Group 1995; Dib et al., 1996; Cox et al., 1990; Hudson et al., 1995; Schuler et al., 1996; Cohen et al., 1993; Chumakov et al., 1995; Hudson et al., 1995); only a few markers were consistently ordered and on all of the maps. By developing the mapping information across the

NBIA1 critical region, we expected to improve marker order and to consolidate all of the marker information from the publicly available maps.

Radiation hybrid mapping is a powerful method for fine-structure mapping of human chromosomes (Goss and Harris, 1975 and 1977ab). A high-resolution radiation hybrid panel was first used to integrate markers in the NBIA1 critical region from the available maps (Stanford TNG3 radiation hybrid panel, Research Genetics, Inc.). Because radiation hybrid mapping is statistical, the data did not suggest a unique order but rather a most likely order. Analysis was performed with the programs RH2PT and RHMINBRK from the RHMAP software package (Boehnke et al., 1991; Boehnke, 1992). RH2PT was first used to assign 37 markers to linkage groups, where each linkage group was assigned a minimum lod score cutoff. The ordering of the linkage groups was determined by framework markers within the groups and by YAC/STS-content mapping data generated with the markers. The RHMINBRK program was then used to order the markers within each linkage group. The resultant marker order was consistent with that determined later when the human BAC DNA pools were screened. The radiation hybrid mapping data was useful as an initial fine-mapping tool to quickly integrate markers from several different maps.

YACs spanning the critical region were also typed with the same markers as above. Since several of the YACs were found to be internally deleted, we decided to screen a BAC DNA library to build a more reliable contig. BAC clones have low frequencies of chimerism and are more stable than YACs or cosmids (Shizuya et al., 1992). BAC maps are not statistically based, can be used to integrate all types of markers, and require simple plus or minus screening by PCR or hybridization. The BAC pools were screened

with over 90 markers; on average, there were 4.5 positive hits per marker tested. Most markers could be ordered just by the number of positive superpool hits they had in common. A second round of screening for some of the positive hits was performed to identify their exact clone addresses. Most of the markers could be placed within eleven small contigs, with ten intervening gaps. These gaps can be closed by additional rounds of end sequencing of the flanking BACs, primer design from the end sequences, and rescreening of the superpools. Identification of additional BAC clones for each marker will increase the depth of the contig. Fingerprinting the clones with restriction endonucleases will confirm contig integrity and help define a minimal tiling path across the region, making it more useful for sequencing efforts. This first-generation BAC/STS-content map of chromosome 20p13-p12.3 provides an essential resource for genetic and molecular studies of this region.

We developed a high-resolution, integrated BAC and YAC contig that spans more than 3 megabases across the NBIA1 critical region of chromosome 20p13-p12.3. There is general consistency between marker order in this map compared to that of the other groups. All of the previously mapped genes from this region, including prodynorphin, could be excluded from the NBIA1 critical region, except for sialoadhesin. Most of the uncharacterized ESTs were excluded as well. With improved mapping information, we found that markers D20S473 and D20S867 were on two overlapping BAC clones, 120C11 (180 kb) and 13N6 (220 kb), respectively. This suggests that the size of the NBIA1 critical region is less than 400 kb.

New families with Hallervorden-Spatz syndrome continue to be collected. These families are being haplotyped with markers from the critical region. Additional obligate

recombinant events in linked, classical HSS families may help to refine the critical region. There are at least two families with classical HSS (HS9, HS50) that do not show linkage to the chromosome 20p13-p12.3 region. Further evidence for locus heterogeneity in HSS may also be observed.

Development of new markers within the 400-kb region may also help to narrow the Hallervorden-Spatz syndrome gene critical region. Several families are homozygous for either marker D20S473 or marker D20S867, but not both (i.e., HS8 and HS12). Perhaps these families are homozygous by chance due to common alleles in the population. Alternatively, the parents may be distantly related. In this case, additional markers may help to narrow the NBIA1 gene region by defining a smaller region of homozygosity between these two markers. Several polymorphic microsatellite markers have already been developed, using the pCR-Blunt vector method (Invitrogen), by our collaborators at UCSF. These markers are from BACs in the NBIA1 critical region.

Special attention is being given to new trinucleotide repeat markers developed in the NBIA1 critical region. With the cloning of the Friedreich's Ataxia gene, frataxin (Campuzano et al., 1996), it has been shown that autosomal recessive disorders can also be caused by triplet repeat expansions (TREs). Human frataxin is a mitochondrial protein (Babcock et al., 1997), and its yeast homologue (YFH1) encodes a mitochondrial protein involved in iron homeostasis and respiratory function. Previously described TRE disorders, including fragile X syndrome and myotonic dystrophy, are either X-linked or autosomal dominant traits. All of the TRE disorders described so far are neurodegenerative diseases. Hallervorden-Spatz syndrome makes an interesting candidate for a second autosomal recessive TRE disorder for two reasons: it is a neurodegenerative

disorder and it is associated with abnormal brain iron accumulation.

BACs from the NBIA1 critical region are being used to screen brain and basal ganglia cDNA libraries, in hopes of identifying unique coding regions in the cloned genomic DNA. If any unique cDNA transcripts are found, they will be characterized to determine whether they correlate with the HSS expression pattern. One method to do this involves hybridizing whole genomic inserts onto a cDNA library using a preannealing procedure to prevent cross hybridization by repetitive DNA sequences (Kendall et al., 1990; Elvin et al., 1990). One of the advantages of this method is that it is possible identify the maximum number of coding regions, including non CpG-island associated genes, with a minimum number of probings of the cDNA library. There are also several disadvantages with this method: expression of the coding sequence of interest may be restricted to a limited number of tissue types, expression may only occur at a specific stage of development, the mRNA species may have low stability or is of very low abundance in the library being screened. Another method, that may alleviate some of these problems, is the direct selection of amplified primary cDNAs using genomic contigs (Morgan et al., 1992). Exon trapping (Church et al., 1994) may also be utilized to identify unique expressed sequences or exons directly from the BACs. It can be used to isolate exons from genomic DNA without the knowledge of the development time or place of gene expression. Exon trapping identifies exon sequences based on splicing signals that flank genomic intron-exon boundaries. This method is applicable for identification of all genes with at least one intron. Exon trapping can be confounded by the identification of exons from genes that are irrelevant to the disease process and by the identification of exon splice sites for which no corresponding transcripts can be found in

any tissue.

Once a putative NBIA1 gene is identified, several forms of analysis may be undertaken in order to confirm that when mutated, it is the disease-causing gene. Identification of open reading frames or clusters of hypomethylated CpG residues in the 5'-end of the gene, and detection of cross-species homology by hybridization to DNA of a wide range of species ('zoo blotting'), suggest a cloned segment of DNA contains a gene. The detection of RNA transcripts by hybridization to Northern blots, followed by the subsequent isolation and sequencing of cDNA clones, is proof that the candidate sequence encodes a transcribed gene. Potential mutations must explain the disease phenotype, be detected in both patients and obligate carriers, and not be homozygous in unaffected individuals. There should not be any recombination between the HSS families and polymorphic probes from within the candidate gene sequence. The gene should be also be expressed in tissues most commonly affected by the disease, and should show some alteration or absence of gene product in affected individuals. This alteration in gene product can be determined by expression of the mutant gene in COS cells. Correlation of the structural and functional data, such as in the creation of a disease model in a transgenic animal, will be the ultimate proof that the NBIA1 disease gene has been found.

The eventual cloning of the Hallervorden-Spatz syndrome gene is important for many reasons. Creation of a transgenic animal model for HSS will expedite study of the underlying biochemical and physiological phenotype of this and related diseases. Methods to identify and diagnose carriers for this lethal condition can be developed. Rational therapies for the management and cure of HSS may be possible with a clearer understanding of the molecular basis and fundamental structural properties of the gene,

and its protein product. Many of the features of HSS, such as retinitis pigmentosa, brain iron accumulation, and dystonia, commonly occur in other isolated disorders. Discovering the processes by which these various events take place in HSS may lead to the identification of common mechanisms for these features in other disorders.

This work is a significant step in identification of the underlying cause of Hallervorden-Spatz syndrome. Using just one large, Amish family, we mapped the first major gene, NBIA1, for HSS to chromosome 20p13-p12.3. By typing additional families, we verified this linkage and gave supporting evidence that there may be at least one other gene responsible for HSS. These additional families also helped to reduce the critical region to that between two very closely linked markers, D20S473 and D20S867. Construction of a physical map of the NBIA1 critical region showed that these two markers were contained on two overlapping BACs, with a total length of approximately 400 kb. Using these two BACs, 120C11 and 13N6, it is now possible to start identifying new candidate genes for HSS. With the collaboration of additional researchers on this project, the first gene responsible for Hallervorden-Spatz syndrome may be identified shortly. The physical map we constructed should be a useful tool for other researchers and groups interested in this area of chromosome 20.

REFERENCES

Aksentijevich I, Pras E, Gruberg L, Shen Y, Holman K, Helling S, Prosen L, Sutherland GR, Richards RI, Ramsburg M, et al.: Refined mapping of the gene causing familial Mediterranean fever, by linkage and homozygosity studies. *Am J Hum Genet* 53:451-461, 1993

Alvarez RA, Ghalayini AJ, Xu P, Hardcastle A, Bhattacharya S, Rao PN, Pettenati MJ, Anderson RE, Baehr W: cDNA sequence and gene locus of the human retinal phosphoinositide-specific phospholipase-Cβ4 (PLCB4). *Genomics* 29:53-61, 1995

Antonarakis SE: Genome linkage scanning: systematic or intelligent? *Nat Genet* 8:211-212, 1994

Babcock M, de Silva D, Oaks R, Davis-Kaplan S, Jiralerspong S, Montermini L, Pandolfo M, Kaplan J: Regulation of mitochondrial iron accumulation by Yfh1p, a putative homolog of frataxin. *Science* 276:1709-1712, 1997

Barrett JH: Genetic mapping based on radiation hybrid data. *Genomics* 13:95-103, 1992

Ben Hamida C, Doerflinger N, Belal S, Linder C, Reutenauer L, Dib C, Gyapay G, Vignal A, Le Paslier D, Cohen D, Pandolfo M, Mokini V, Novelli G, Hentati F, Ben Hamida M, Mandel J-L, Koenig M: Localization of Friedreich ataxia phenotype with selective vitamin E deficiency to chromosome 8q by homozygosity mapping. *Nat Genet* 5:195-200, 1993

Ben-Shachar D, Livne E, Spanier I, Leenders KL, Youdim MB: Typical and atypical neuroleptics induce alteration in blood-brain and brain 59$FeCl_3$ uptake. *J Neurochem* 62:1112-1118, 1994

Bishop DT, Crockford GP: Comparisons of radiation hybrid mapping and linkage mapping. *Cytogenet Cell Genet* 59:93-95, 1992

Boehnke M: Multipoint analysis for radiation hybrid mapping. *Ann Med* 24:383-386, 1992

Boehnke M: Radiation hybrid mapping by minimization of the number of obligate chromosome breaks. *Cytogenet Cell Genet* 59:96-98, 1992

Boehnke M, Lange K, Cox DR: Statistical methods for multipoint radiation hybrid mapping. *Am J Hum Genet* 49:1174-1188, 1991

Boehnke M, Lunetta K, Hauser E, Lange K, Uro J, VanderStoep J: RHMAP: Statistical Package for Multipoint Radiation Version 3.0, September 1996

Botstein D, White RL, Skolnick M, Davis RW: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331, 1980

Burke DT, Carle GF, Olson MV: Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236:806-812, 1987

Campuzano V, Montermini L, Moltò MD, Pianese L, Cossée M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Cañizares C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R, Palau F, Patel PI, Di Donato S, Mandel J-L, Cocozza S, Koenig M, Pandolfo M: Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271:1423-1427, 1996

Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH: Nonuniform recombination within the human beta-globin gene cluster. *Am J Hum Genet* 36:1239-1258, 1984

Chumakov IM, Rigault P, Le Gall I, Bellanné-Chantelot C, Billault A, Guillou S, Soularue P, Guasconi G, Poullier E, Gros I, Belova M, Sambucy J-L, Susini L, Gervy P, Glibert F, Beaufils S, Bui H, Massart C, De Tand M-F, Dukasz F, Lecoulant S, Ougen P, Perrot V, Saumier M, Soravito C, Bahouayila R, Cohen-Akenine A, Barillot E, Bertrand S, Codani J-J, Caterina D, Georges I, Lacroix B, Lucotte G, Sahbatou M, Schmit C, Sangouard M, Tubacher E, Dib C, Fauré S, Fizames C, Gyapay G, Millasseau P, Nguyen S, Muselet D, Vignal A, Morissette J, Menninger J, Lieman J, Desai T, Banks A, Bray-Ward P, Ward D, Hudson TJ, Gerety SS, Foote S, Stein LD, Page DC, Lander ES, Weissenbach J, Le Paslier D, Cohen D: A YAC contig map of the human genome. *Nature* 377:174-297, 1995

Church DM, Stotler CJ, Rutter JL, Murrell JR, Trofatter JA, Buckler AJ: Isolation of genes from complex sources of mammalian DNA using exon amplification. *Nat Genet* 6:98-105, 1994

Church G, Kieffer-Higgins S: Multiplex DNA sequencing. *Science* 240:185-188, 1988

Cohen D, Chumakov I, Weissenbach J: A first-generation physical map of the human genome. *Nature* 366:698-701, 1993

Connor JR, Fine RE: The distribution of transferrin immunoreactivity in the rat central nervous system. *Brain Res* 368:319-328, 1986

Connor JR, Phillips TM, Lakshman MR, Barron KD, Fine RE, Csiza CK: Regional variation in the levels of transferrin in the CNS of normal and myelin-deficient rats. *J Neurochem* 49:1523-1529, 1987

Connor JR, Snyder BS, Beard JL, Fine RE, Mufson EJ: Regional distribution of iron and iron regulatory proteins in the brain in aging and Alzheimer's disease. *J Neurosci Res* 31:327-335, 1992

Cottingham RW, Jr., Idury RM, Schäffer AA: Faster sequential genetic linkage computations. *Am J Hum Genet* 53:252-263, 1993

Cox DR, Burmeister M, Price ER, Kim S, Myers RM: Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* 250:245-250, 1990

Current Protocols in Human Genetics (Dracopoli NC, Haines JL, Korf BR, Moir DT, Morton CC, Seidman CE, Seidman JG, Smith DR and Boyle AL, eds.). Greene Publishing and John Wiley & Sons, New York, 1994

Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R: Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 6:575-577, 1990

Dedman DT, Treffry A, Candy JM, Taylor GA, Morris CM, Bloxham CA, Perry RH, Edwardson JA, Harrison PM: Iron and aluminum in relation to brain ferritin in normal individuals and Alzheimer's disease and chronic renal-dialysis patients. *Biochem J* 287:509-514, 1992

Defendini R, Markesbery WR, Mastri AR, Duffy PE: Hallervorden-Spatz disease and infantile neuroaxonal dystrophy: Ultrastructural observations, anatomical pathology, and nosology. *J Neurosci* 20:7-23, 1973

Dexter DT, Carayon A, Javoy-Agid F, Agid Y, Wells FR, Daniel SE, Lees AJ, Jenner P, Marsden CD: Alterations in the levels of iron, ferritin and other trace metals in Parkinson's disease and other neurodegenerative diseases affecting basal ganglia. *Brain* 114:1953-1975, 1991

Dexter DT, Wells FR, Lees AJ, Agid F, Agid Y, Jenner P, Marsden CD: Increased nigral iron content and alteration in other metals occurring in Parkinson's disease. *J Neurochem* 52:1830-1836, 1989

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, J. W: A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152-154, 1996

DiLella AG, Hawkins A, Craig RJ, Schreiber SL, Griffin CA: Chromosomal band assignments of the genes encoding human FKBP12 and FKBP13. *Biochem Biophys Res Comm* 189:819-823, 1992

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS: 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* 19:4008, 1991

Dooling EC, Schoene WC, Richardson EP, Jr.: Hallervorden-Spatz syndrome. *Arch Neurol* 30:70-83, 1974

Drayer B: Magnetic resonance imaging and brain iron: implications in the diagnosis and pathochemistry of movement disorders and dementia. *Barrow Neurol Inst Q* 3:15-30, 1987

Edwards JH: Exclusion mapping. *J Med Genet* 24:539-543, 1987

Edwards A, Civitello A, Hammond HA, Caskey CT: DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 49:746-756, 1991

Ehinger B, Narfstrom K, Nilsson SE, van Veen T: Photoreceptor degeneration and loss of immunoreactive GABA in the Abyssinian cat retina. *Exp Eye Res* 52:17-25, 1991

Elvin P, Slynn G, Black D, Graham A, Butler R, Riley J, Anand R, Markham AF: Isolation of cDNA clones using yeast artificial chromosome probes. *Nucleic Acids Res* 18:3913-3917, 1990

Esbenshade TA, Hirasawa A, Tsujimoto G, Tanaka T, Yano J, Minneman KP, Murphy TJ: Cloning of the human a1d-adrenergic receptor and inducible expression of three human subtypes in SK-N-MC cells. *Mol Pharmacol* 47:977-985, 1995

Farber DB, Lolley R N: Cyclic guanosine monophosphate: elevation in degenerating photoreceptor cells of the C3H mouse retina. *Science* 186: 449-451, 1974

Farber DB, Lolley RN: Enzyme basis for cyclic GMP accumulation in degenerative photoreceptor cells of mouse retina. *J. Cyclic Nucleotide Res.* 2: 139-148, 1976

Farrall M: Homozygosity mapping: familiarity breeds debility. *Nat Genet* 5:107-108, 1993

Gallucci M, Cardona F, Arachi M, Splendiani A, Bozzao A, Passariello R: Follow-up MR studies in Hallervorden-Spatz disease. *J Comput Tomogr* 14:118-120, 1990

Goldrick MM, Kimball GR, Liu Q, Martin LA, Sommer SS, Tseng JY-H: NIRCA: a rapid robust method for screening for unknown point mutations. *BioTechniques* 21:106-112, 1996

Goodman L: Alzheimer's disease: a clinicopathologic analysis of twenty-three cases with a theory of pathogenesis. *J Nerv Ment Dis* 118:97-130, 1953

Gopal Rao VVN, Loffler C, Battey J, Hansmann I: The human gene for oxytocin-neurophysin I (OXT) is physically mapped to chromosome 20p13 by in situ hybridization. *Cytogenet Cell Genet* 61:271-273, 1992

Gordon J: Julius Hallervorden. *Neurology* 43:1452, 1993

Goss SJ, Harris H: New method for mapping genes in human chromosomes. *Nature* 255:680-684, 1975

Goss SJ, Harris H: Gene transfer by means of cell fusion I. Statistical mapping of the human X-chromosome by analysis of radiation-induced gene segregation. *J Cell Sci* 25:17-37, 1977

Goss SJ, Harris H: Gene transfer by means of cell fusion. II. The mapping of 8 loci on human chromosome 1 by statistical analysis of gene assortment in somatic cell hybrids. *J Cell Sci* 25:39-57, 1977b

Green ED, Green P: Sequence-tagged site (STS) content mapping of human chromosomes: theoretical considerations and early experiences. *PCR Methods Appl* 1:77-90, 1991

Green ED, Olson MV: Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: A model for human genome mapping. *Science* 250:94-98, 1990

Green ED, Olson MV: Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc Natl Acad Sci U S A* 87:1213-1217, 1990

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J: The 1993-94 Généthon human genetic linkage map. *Nat Genet* 7:246-339, 1994

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme J-F, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN: A radiation hybrid map of the human genome. *Hum Mol Genet* 5:339-346, 1996

Haldane JBS, Smith CAB: A new estimate of the linkage between the gene for colour-blindness and hemophilia in man. *Ann Eugenics* 14:10-31, 1947

Hallervorden J: Uber eine familiare Erkrankung im extrapyramidalen System. *Dtsch Z Nervenheilkd* 81:204-210, 1924

Hallervorden J, Spatz H: Eigenartige Erkrankung im extrapyramidalen System mit besonderer Beteiligung des Globus pallidus und der Substantia nigra Ein Beitrag zu den Beziehungen zwischen diesen beiden Zentren. *Z Ges Neurol Psychiat* 79:254-302, 1922

Halliwell B, Gutteridge JMC: Oxygen toxicity, oxygen radicals, transition metals and disease. *Biochem J* 219:1-14, 1984

Halliwell B, Gutteridge JMC: The importance of free radicals and catalytic metal ions in human disease. *Mol Aspects Med* 8:89-193, 1985

Hardy G, Weinberg W: The Hardy-Weinberg Law of Genetic Equilibrium. 1908

Haseman JK, Elston RC: The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3-19, 1972

Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander ES: Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 2:204-211, 1992

Hayflick SJ, Taylor T, McKinnon W, Guttmacher AE, Litt M, Zonana J: Clouston syndrome (hidrotic ectodermal dysplasia) is not linked to keratin gene clusters on chromosomes 12 and 17. *J Invest Dermatol* 107:11-14, 1996

Hentati A, Pericak-Vance MA, Hung W-Y, Belal S, Laing N, Boustany R-M, Hentati F, Ben Hamida M, Siddique T: Linkage of 'pure' autosomal recessive familial spastic paraplegia to chromosome 8 markers and evidence of genetic locus heterogeneity. *Hum Mol Genet* 3:1263-1267, 1994

Higgins JJ, Patterson MC, Papadopoulos NM, Brady RO, Pentchev PG, Barton NW: Hypoprebetalipoproteinemia, acanthocytosis, retinitis pigmentosa, and pallidal degeneration (HARP syndrome). *Neurology* 42:194-198, 1992

Hill JM: Iron concentration reduced in ventral pallidum, globus pallidus, and substantia nigra by GABA-transaminase inhibitor, gamma-vinyl GABA. *Brain Res* 342:18-25, 1985

Hill JM, Switzer III RC: The regional distribution and cellular localization of iron in the rat brain. *Neuroscience* 11:595-603, 1984

Hill WG, Robertson A: The effects of inbreeding at loci with heterozygote advantage. *Genetics* 60:615-628, 1968

Hirsch EC, Brandel JP, Galle P, Javoy-Agid F, Agid Y: Iron and aluminum increase in the substantia nigra of patients with Parkinson's disease: an X-ray microanalysis. *J Neurochem* 56:446-451, 1991

Hoffman CS: Preparation of yeast DNA. *In* Current Protocols in Human Genetics (Dracopoli NC, Haines JL, Korf BR, Moir DT, Morton CC, Seidman CE, Seidman JG, Smith DR and Boyle AL, eds.) pp. 5.9.23-5.9.25. (Greene Publishing and John Wiley & Sons, New York), 1993

Houwen RHJ, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, Sandkuijl LA, Freimer NB: Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nat Genet* 8:380-386, 1994

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu S-H, Hu X, Colbert AME, Rosenberg C, Reeve-Daly MP, Rozen S, Hui L, Wu X, Vestergaard C, Wilson KM, Bae JS, Maitra S, Ganiatsas S, Evans CA, DeAngelis MM, Ingalls KA, Nahf RW, Horton LT, Jr., Anderson MO, Collymore AJ, Ye W, Kouyoumjian V, Zemsteva IS, Tam J, Devine R, Courtney DF, Renaud MT, Nguyen H, O'Connor TJ, Fizames C, Fauré S, Gyapay G, Dib C, Morissette J, Orlin JB, Birren BW, Goodman N, Weissenbach J, Hawkins TL, Foote S, Page DC, Lander ES: An STS-based map of the human genome. *Science* 270:1945-1954, 1995

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ: A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet* 6:84-89, 1994

Jackson AJ, Michael LM, Schumacher HJ: Improved tissue solubilization for atomic absorption. *Anal Chem* 44:1064-1065, 1972

Jirik FR, Anderson LL, Duncan AMV: The human protein-tyrosine phosphatase PTP-alpha/LRP gene (PTPA) is assigned to chromosome 20p13. *Cytogenet Cell Genet* 60:117-118, 1992

Kelm S, Schauer R, Crocker PR: The sialoadhesins - a family of sialic acid-dependent cellular recognition molecules within the immunoglobulin superfamily. *Glycoconjugate J* 13:913-926, 1996

Kendall E, Sargent CA, Campbell RD: Human major histocompatibility complex contains a new cluster of genes between the HLA-D and complement C4 loci. *Nucleic Acids Res* 18:7251-7257, 1990

Kibar Z, Der Kaloustian VM, Brais B, Hani V, Fraser FC, Rouleau GA: The gene responsible for Clouston hidrotic ectodermal dysplasia maps to the pericentromeric region of chromosome 13q. *Hum Mol Genet* 5:543-547, 1996

Kim U-J, Birren BW, Slepak T, Mancino V, Boysen C, Kang H-L, Simon MI, Shizuya H: Construction and characterization of a human bacterial artificial chromosome library. *Genomics* 34:213-218, 1996

Kornyey S: Die Stoffwechechselstorungen bei der Hallervorden-Spatzschen Krankheit. *Z Neurol* 205:178, 1964

Kwitek-Black AE, Carmi R, Duyk GM, Buetow KH, Elbedour K, Parvari R, Yandava CN, Stone EM, Sheffield VC: Linkage of a Bardet-Biedl syndrome to chromosome 16q and evidence for non-allelic genetic heterogeneity. *Nat Genet* 5:392-396, 1993

Lai YL, Jacoby RO, Jensen JT, Yao PC: Retinitis pigmentosa. *Am J Path* 98:281-284, 1980

Lam E, Martin MM, Timerman AP, Sabers C, Fleischer S, Lukas T, Abraham RT, O'Keefe SJ, O'Neill EA, Wiederrecht GJ: A novel FK506 binding protein can mediate the immunosuppressive effects of FK506 and is associated with the cardiac ryanodine receptor. *J Biol Chem* 270:26511-26522, 1995

Lander ES, Botstein D: Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236:1567-1570, 1987

Lange K, Boehnke M, Cox DR, Lunetta KL: Statistical methods for polyploid radiation hybrid mapping. *Genome Res* 5:136-150, 1995

Lathrop GM, Lalouel JM, Julier C, Ott J: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A* 81:3443-3446, 1984

LeVine SM, Macklin WB: Iron-enriched oligodendrocytes: a reexamination of their spatial distribution. *J Neurosci Res* 26:508-512, 1990

Litt M, Buroker NE, Kondoleon S, Douglass J, Liston D, Sheehy R, Magenis RE: Chromosomal localization of the human proenkephalin and prodynorphin genes. *Am J Hum Genet* 42:327-334, 1988

Litt M, Hauge XY, Sharma V: Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. *BioTechniques* 15:280-284, 1993

Litt M, Luty JA: A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44:397-401, 1989

Louie LG, King MC: A novel approach to establishing permanent lymphoblastoid cell lines: Epstein-Barr virus transformation of cryopreserved lymphocytes. *Am J Hum Genet* 48:637-638, 1991

Luckenbach MW, Green WR, Miller NR, Moser HW, Clark AW, Tennekoon G: Ocular clinicopathologic correlation of Hallervorden-Spatz syndrome with acanthocytosis and pigmentary retinopathy. *Am J Ophthalmol* 95:369-382, 1983

Michaelis EK, Belieu RM, Grubbs RD, Michaelis ML, Chang HH: Differential effects of metal ligands on synaptic membrane glutamate binding uptake systems. *Neurochem Res* 7:423-436, 1982

Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215, 1988

Morgan JG, Dolganov GM, Robbins SE, Hinton LM, Lovett M: The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucleic Acids Res* 20:5173-5179, 1992

Morton N: Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277-318, 1955

Mucklow S, Hartnell A, Mattei M-G, Gordon S, Crocker PR: Sialoadhesin (*Sn*) maps to mouse chromosome 2 and human chromosome 20 and is not linked to the other members of the sialoadhesin family, CD22, MAG, and CD33. *Genomics* 28:344-346, 1995

Mullis KB, Faloona FA: Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155:335-350, 1987

Murray JC, Buetow KH, Weber JL, Ludwigsen S, Scherpier-Heddema T, Manion F, Quillen J, Sheffield V, Sunden S, Duyk GM, Weissenbach J, Gyapay G, Dib C, Morissette J, Lathrop GM, Vignal A, White R, Matsunami N, Gerken S, Melis R, Albertsen H, Plaetke R, Odelberg S, Ward D, Dausset J, Cohen D, Cann H: A comprehensive human linkage map with centimorgan density. *Science* 265:2049-2054, 1994

Mutoh K, Okuno T, Ito M: MR imaging of a group I case of Hallervorden-Spatz disease. *J Comput Tomogr* 12:851-853, 1988

Newell FW, Johnson RO, II, Huttenlocher PR: Pigmentary degeneration of the retina in the Hallervorden-Spatz syndrome. *Am J Ophthalmol* 88:467-471, 1979

NIH/CEPH Collaborative Mapping Group: A comprehensive genetic linkage map of the human genome. *Science* 258:67-86, 1992

Olson M, Hood L, Cantor C, Botstein D: A common language for physical mapping of the human genome. *Science* 245:1434-1435, 1989

Online Mendelian Inheritance in Man, OMIM (TM). Center for Medical Genetics, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 1997.
World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/

Ott J: Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci USA* 86:4175-4178, 1989

Ott J: Strategies for characterizing highly polymorphic markers in human gene mapping. *Am J Hum Genet* 51:283-290, 1992

Ozelius LJ, Kramer PL, deLeon D, Risch N, Bressman SB, Schuback DE, Brin MF, Kwiatkowski DJ, Burke RE, Gusella JF, Fahn S, Breakefield X: Strong allelic association between the torsion dystonia gene (DYT1) and loci on chromosome 9q34 in Ashkenazi Jews. *Am J Hum Genet* 50:619-628, 1992

Park BE, Netsky MG, Betsill WL: Pathogenesis of pigment and spheroid formation in Hallervorden-Spatz syndrome and related disorders. *Neurology* 25:1172-1178, 1975

Perry TL, Norman MG, Yong VW, Whiting S, Crichton JU, Hansen S, Kish SJ: Hallervorden-Spatz disease: cysteine accumulation and cysteine dioxygenase deficiency in the globus pallidus. *Ann Neurol* 18:482-489, 1985

Puffenberger EG, Kauffman ER, Bolk S, Matise TC, Washington SS, Angrist M, Weissenbach J, Garver KL, Mascari M, Ladda R, Slaugenhaupt SA, Chakravarti A: Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. *Hum Mol Genet* 3:1217-1225, 1994

Rapp LM, Wiegand RD, Anderson RE: Ferrous iron-mediated retinal degeneration: role of rod outer segment lipid peroxidation. 1977

Rick JT, Tunnicliff G, Kerkut GA, Fulker DW, Wilcock J, Broadhurst PL: GAG production in brain cortex related to activity and avoidance behaviour in eight strains of rat. *Brain Research* 32:234-238, 1971

Riederer P, Sofic E, Rausch WD, Schmidt B, Reynolds GP, Jellinger K, Youdim MB: Transition metals, ferritin, glutathione and ascorbic acid in parkinsonian brains. *J Neurochem* 52:515-520, 1989

Roth A, et al.: Pigmentary retinal dystrophy in Hallervorden-Spatz disease: clinicopathological report of a case. *Surv Ophthalmol* 16:24-35, 1971

Schäffer AA: Faster linkage analysis computations for pedigrees with loops or unused alleles. *Hum Hered* 46:226-235, 1996

Schäffer AA, Gupta SK, Shriram K, Cottingham RW, Jr.: Avoiding recomputation in linkage analysis. *Hum Genet* 44:225-237, 1994

Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tomé P, Aggarwal A, Bajorek E, Bentolila S, Birren BB, Butler A, Castle AB, Chiannilkulchai N, Chu A, Clee C, Cowles S, Day PJR, Dibling T, Drouot N, Dunham I, Duprat S, East C, Edwards C, Fan J-B, Fang N, Fizames C, Garrett C, Green L, Hadley D, Harris M, Harrison P, Brady S, Hicks A, Holloway E, Hui L, Hussain S, Louis-Dit-Sully C, Ma J, MacGilvery A, Mader C, Maratukulam A, Matise TC, McKusick KB, Morissette J, Mungall A, Muselet D, Nusbaum HC, Page DC, Peck A, Perkins S, Piercy M, Qin F, Quackenbush J, Ranby S, Reif T, Rozen S, Sanders C, She X, Silva J, Slonim DK, Soderlund C, Sun W-L, Tabar P, Thangrajah T, Vega-Czarny N, Vollrath D, Voyticky S, Wilmer T, Wu X, Adams MD, Auffray C, Walter NAR, Brandon R, Dehejia A, Goodfellow PN, Houlgatte R, Hudson JR, Jr., Ide SE, Iorio KR, Lee WY, Seki N, Nagase T, Ishikawa K, Nomura N, Phillips C, Polymeropoulos MH, Sandusky M, Schmitt K, Berry R, Swanson K, Torres R, Venter JC, Sikela JM, Beckmann JS, Weissenbach J, Myers RM, Cox DR, James MR, et al.: A gene map of the human genome. *Science* 274:540-546, 1996

Sethi KD, Adams RJ, Loring DW, El Gammal T: Hallervorden-Spatz syndrome: clinical and magnetic resonance imaging correlations. *Ann Neurol* 24:692-694, 1988

Sheffield VC, Carmi R, Kwitek-Black A, Rokhlina T, Nishimura D, Duyk GM, Elbedour K, Sunden SL, Stone EM: Identification of a Bardet-Biedl syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. *Hum Mol Genet* 3:1331-1335, 1994

Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, Gastier JM, Pulido JC, Yandava C, Sunden SL, Mattes G, Businga T, McClain A, Beck J, Scherpier T, Gilliam J, Zhong J, Duyk GM: A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. *Hum Mol Genet* 4:1837-1844, 1995

Shevell M: Racial hygiene, active euthanasia, and Julius Hallervorden. *Neurology* 42:2214-2219, 1992

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M: Cloning and stable maintenance of 300-kilobase pair fragments of human DNA in E coli using an F-factor-based vector. *Proc Natl Acad Sci U S A* 89:8794-8797, 1992

Shoham S, Wertman E, Ebstein RP: Iron accumulation in the rat basal ganglia after excitatory amino acid injections-dissociation from neuronal loss. *Exp Neurol* 118:227-241, 1992

Sofic E, Paulus W, Jellinger K, Riederer P, Youdim MB: Selective increase of iron in the substantia nigra zona compacta of parkinsonian brains. *J Neurochem* 56:978-982, 1991

Sofic E, Riederer P, Heinsen H, Beckmann H, Reynolds GP, Hebenstreit G, Youdim MB: Increased iron (III) and total iron content in post mortem substantia nigra of parkinsonian brain. *J Neural Transm* 74:199-205, 1988

Spatz H: Uber des Eisenmachweiss im Gehirn besonders in Zentren des extrapyramidalmotorischen Systems. *Z Ges Neurol Psychiat* 77:261, 1922

Spatz H: Uber stoffwechseleigent. *Z Ges Neurol Psychiat* 78:641, 1922

Stearman R, Yuan DS, Yamaguchi-Iwai Y, Klausner RD, Dancis A: A permease-oxidase complex involved in high-affinity iron uptake in yeast. *Science* 271:1552-1557, 1996

Strautnieks SS, Thompson RJ, Hanukoglu A, Dillon MJ, Hanukoglu I, Kuhnle U, Seckl J, Gardiner RM, Chung E: Localisation of pseudohypoaldosteronism genes to chromosome 16p12.2-13.11 and 12p13.1-pter by homozygosity mapping. *Hum Mol Genet* 5:293-299, 1996

Summar ML, Phillips JA, III B, J., Castiglione CM, Kidd KK, Maness KJ, Weiffenbach B, Gravius TC: Linkage relationships of human arginine vasopressin-neurophysin-II and oxytocin-neurophysin-I to prodynorphin and other loci on chromosome 20. *Mol Endocrinol* 4:947-950, 1990

Swaiman KF: Hallervorden-Spatz syndrome and brain iron metabolism. *Arch Neurol* 48:1285-1293, 1991

Swaiman KF, Smith SA, Trock GL, Siddiqui AR: Sea-blue histiocytes, lymphocytic cytosomes, movement disorder and 59Fe-uptake in basal ganglia: Hallervorden-Spatz disease or ceroid storage disease with abnormal isotope scan? *Neurology* 33:301-305, 1983

Swisher CN, et al.: Coexistence of Hallervorden-Spatz disease with acanthocytosis. *Trans Am Neurol Assoc* 97:212, 1972

Szanto J, Gallyas F: A study of iron metabolism in neuropsychiatric patients: Hallervorden-Spatz disease. *Arch Neurol* 14:438-442, 1966

Tanfani G, Mascalchi M, Dal Pozzo GC, Taverni N, Saia A, Trevisan C: MR imaging in a case of Hallervorden-Spatz disease. *J Comput Tomogr* 11:1057-1058, 1987

Taylor TD, Hayflick SJ, McKinnon W, Guttmacher AE, Hovnanian A, Litt M, Zonana J: Confirmation of linkage of Clouston syndrome (hidrotic ectodermal dysplasia) to 13q11-q12.1 with evidence for multiple independent mutations. *J Invest Dermatol* 111:83-85, 1998

Taylor TD, Litt M, Kramer P, Pandolfo M, Angelini L, Nardocci N, Davis S, Pineda M, Hattori H, Flett PJ, Cilio MR, Bertini E, Hayflick SJ: Homozygosity mapping of Hallervorden-Spatz syndrome to chromosome 20p12.3-p13. *Nat Genet* 14:479-481, 1996

Taylor TD, Litt M, Kramer P, Pandolfo M, Angelini L, Nardocci N, Davis S, Pineda M, Hattori H, Flett PJ, Cilio MR, Bertini E, Hayflick SJ: Homozygosity mapping of Hallervorden-Spatz syndrome to chromosome 20p12.3-p13. *Nat Genet* 16:109, 1997

Terwilliger JD: A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am J Hum Genet* 56:777-787, 1995

Terwilliger JD, Ott J: Handbook of Human Genetic Linkage (Johns Hopkins University Press, Baltimore, 1994)

Thompson EA: Crossover counts and likelihood in multipoint linkage analysis. *IMA J Math Appl Med Biol* 4:93-108, 1987

Tripathi RC, Tripathi BJ, Bauserman SC, Park JK: Clinicopatholgic correlation and pathogenesis of ocular and central nervous system manifestations in Hallervorden-Spatz syndrome. *Acta Neuropathol (Berl)* 83:113-119, 1992

Utah Marker Development Group: A collection of ordered tetranucleotide-repeat markers from the human genome. *Am J Hum Genet* 57:619-628, 1995

Vakili S, Drew AL, Von Schuching S, Becker D, Zeman W: Hallervorden-Spatz syndrome. *Arch Neurol* 34:729-738, 1977

Vignal A, Gyapay G, Hazan J, Nguyen S, Dupraz C, Cheron N, Becuwe N, Trachant M, Weissenbach J: A nonradioactive multiplex procedure for genotyping of microsatellite markers. *In* Methods in Molecular Genetics, Vol. 1. Gene and Chromosome Analysis: Part A. (KW Adolph, ed.) pp. 211-221. (Academic Press, San Diego), 1993

Volkl A, Ule G: Trace elements in human brain: Iron concentrations of 13 brain areas as a function of age. *Z Neurol* 202:331-338, 1972

Wald A: Sequential Analysis. (John Wiley & Sons, New York), 1947

Weber J: Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. *Genomics* 7:524-530, 1990

Weber JL, May PE: Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44:388-396, 1989

Weeks DE, Lange K: The affected-pedigree-member method of linkage analysis. *Am J Hum Genet* 42:315-326, 1988

Weeks DE, Lehner T, Ott J: Preliminary ranking procedures for multilocus ordering based on radiation hybrid data. *Cytogenet Cell Genet* 59:125-127, 1992

Weeks DE, Ott J, Lathrop GM: SLINK: A general simulation program for linkage analysis. *Am J Hum Genet* 47 (Suppl.):A204, 1990

Weeks DE, Sobel E, O'Connell JR, Lange K: Computer programs for multilocus haplotyping of general pedigrees. *Am J Hum Genet* 56:1506-1507, 1995

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M: A second-generation linkage map of the human genome. *Nature* 359:794-801, 1992

Wunderle V, Dib C, Fizames C, Morissette J, Hazan J, Hansmann I, Whitehouse D, Vergnaud G, Weissenbach J: The EUROGEN map of human chromosome 20. *Eur J Hum Genet* 2:242-243, 1994

Yang-Feng TL, Han H, Lomasney JW, Caron MG: Localization of the cDNA for an a a_1-adrenergic receptor subtype (ADRA1D) to chromosome band 20p13. *Cytogenet Cell Genet* 66:170-171, 1994

APPENDICES

PROTOCOLS

OTHER PAPERS

CLOUSTON SYNDROME (HIDROTIC ECTODERMAL DYSPLASIA) IN NOT LINKED TO KERATIN GENE CLUSTERS ON CHROMOSOMES 12 AND 17

CONFIRMATION OF LINKAGE OF CLOUSTON SYNDROME (HIDROTIC ECTODERMAL DYSPLASIA) TO 13q11-q12.1 WITH EVIDENCE FOR MULTIPLE INDEPENDENT MUTATIONS

PROTOCOLS

RADIATION HYBRID SCREENING AND ANALYSIS

(High Resolution Stanford TNG3 Radiation Hybrid Panel, Catalog No. RH03.03, Research Genetics, Inc.)

1. Description

This panel of 90 radiation hybrid clones of the whole human genome was created by the Stanford Human Genome Center. The human lymphoblastoid cell line (donor RM) was exposed to 50,000 rad of x-rays and fused with nonirradiated thymidine-kinase deficient hamster recipients cells (A23), creating a panel of independent somatic hybrid clones with 100 kb resolution. Approximately 16% of the human genome is retained in each hybrid. The average size of the human fragments is 800 kb. 1% X-ray breakage equals about 4 kb. The average resolution of this panel is 60 kb. This panel of TNG3 DNA is ready to be used for ordering markers in the region of interest as well as establishing the distance between these markers.

The Stanford TNG3 Radiation Hybrid Panel contains a vial of DNA from each of 90 radiation hybrid clones, plus the two control DNAs (RM donor and A23 recipient). Each vial contains DNA for PCR at a concentration of 25 ng/µl. Most PCR reactions were performed using 25 ng of DNA in 10 µl PCR reactions (see following worksheet for details).

2. PCR Conditions

A. PCR conditions will vary according to primer length and base composition, reaction volume, and the type of temperature cycler used. Research Genetics, Inc. recommends using one 1 µl of DNA per PCR reaction.

- PCR primer concentrations, number of PCR cycles and annealing temperatures may vary for each marker. It is best to work these out with test reactions first. I found that using 0.125 µl for each primer (20 µM) in a 10 µl reaction was usually sufficient. The best combination for ease in scoring the gel is a low primer concentration and a low number of PCR cycles.

B. An automated multi-channel pipette (0.2-10 µl, Rainin Instrument Company) can be used to load DNA into the PCR reaction tubes to minimize the amount of time the plates are left unfrozen. When measuring small quantities it is important to ensure that all pipette tips are snugly in place. It is also much easier to pipette small quantities of DNA into tubes already containing the PCR premix.

C. When analyzing the PCR reactions on a gel (generally 0.5X TBE 2% agarose), it is necessary to have included both genomic and water controls in your PCR set up. A DNA ladder should also be run alongside for size comparison. The automated multi-channel pipette can be used to add samples into the gel in order to minimize loading time.

- By arranging the DNA samples in a staggered format (as seen in the worksheet), an eight-channel pipette can be used to transfer the PCR products to an agarose gel (with six rows of 18 wells each) such that every other well is loaded for each column on the PCR plate. Each row on the gel will comprise of two adjacent columns on the PCR plate. With this method, the gels can be quickly loaded and the clones remain in numerical order. This also leaves room for a size control on both sides of each row. All reactions can be run at the same time on just one gel (just keep a careful watch to make sure one row does not run into another).

3. Analysis

RH mapping can be achieved using a statistical program (RHMAP; Boehnke et al., 1991; Lange et al., 1995; Boehnke et al., 1996) that will provide the best map along with a measure of the relative likelihood of one order versus another. This type of analysis has been shown to successfully generate the order of markers on the RH map that is significantly more likely than in any alternative order.

RH2PT is a program for data description and two-point analysis. It estimates locus specific retention probabilities, pairwise breakage probabilities and two-point lod scores for linkage of the various marker pairs and linkage groups.

RHMINBRK is a program for multilocus ordering by minimization of the number of obligate chromosome breaks.

RHMAXLIK is a program for multilocus ordering by maximization of the likelihood

of the hybrid data under a variety of breakage and retention models.

Both RHMINBRK and RHMAXLIK can evaluate a user-specified list of locus orders. They each employ one of several strategies of combinatorial optimization to attempt to identify the best locus orders: stepwise locus ordering, simulated annealing, and branch and bound. They can identify influential hybrids that have a large impact on ordering conclusions.

Software for the analysis of results is available free of charge via the University of Michigan's Statistical Genetics Software web site (http://www.sph.umich.edu/group/statgen/software) or you may e-mail Dr. Michael Boehnke at boehnke@umich.edu to request a copy.

PCR HEADING: Amplify TNG3 RH Panel with marker: **DATE:** / / **PAGE #:**

Condition: : - °C TD w/30 sec ext @ 72 °C (35X, 37X, 40X) + 94 °C 4' start

DNA Conc (µg/ml): 25 **DNA Vol/Rxn (µl):** 1 **Total Rxn Vol (µl):** 10

PCR REACTION PARAMETERS:

LOCUS/GROUPING:

	µl/rxn	100
sterile D.I. H_2O:	5.4	540.00
1.25 mM dNTPs:	1.6	160.00
10X PCR buffer:	1	100.00
25 mM MgCl:	0.6	60.00
20 mM forward primer:	0.125	12.50
20 mM reverse primer:	0.125	12.50
25 mM spermidine:	0.1	10.00
Taq polymerase (5U/µl):	0.05	5.00
100% DMSO:	0	0.00
TOTAL PREMIX VOLUME (µl):	9	900.00

_____ **Approx. Size (bp)**

1.25 **µl 10x Stop Buffer (xcff only)**

5 **µl 10 mg/ml EtBr in 75 ml gel**

2 **µl 100 bp ladder (50ng/µl)**

Time	
From:	**To:**

_____ **Start setup**

_____ **PCR**

_____ **Gel (100V, max Amp)**

_____ **Hits**

_____ **? hits**

_____ **Entered in RH db**

Run products out on 2% 0.5X TBE agarose gels.

PCR rack set-up (TNG3 Radiation Hybrid Panel):

	1	2	3	4	5	6	7	8	9	10	11	12
A	01	02	17	18	33	34	49	50	65	66	81	82
B	03	04	19	20	35	36	51	52	67	68	83	84
C	05	06	21	22	37	38	53	54	69	70	85	86
D	07	08	23	24	39	40	55	56	71	72	87	88
E	09	10	25	26	41	42	57	58	73	74	89	90
F	11	12	27	28	43	44	59	60	75	76	RM	A23
G	13	14	29	30	45	46	61	62	77	78	genomic	H_2O
H	15	16	31	32	47	48	63	64	79	80		
	Gel A-top		Gel A-middle		Gel A-bottom		Gel B-top		Gel B-middle		Gel B-bottom	

Gel A

top	ladder-01-02-03-04-05-06-07-08-09-10-11-12-13-14-15-16-ladder
middle	ladder-17-18-19-20-21-22-23-24-25-26-27-28-29-30-31-32-ladder
bottom	ladder-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-ladder

Gel B

top	lad.-49-50-51-52-53-54-55-56-57-58-59-60-61-62-63-64-lad.
middle	lad.-65-66-67-68-69-70-71-72-73-74-75-76-77-78-79-80-lad.
bottom	lad.-81-82-83-84-85-86-87-88-89-90-RM-A23-(1-301)-H_2O-lad.

Worksheet 1. RH PCR Setup

BAC POOL SCREENING

(Human Bacterial Artificial Chromosome DNA Pools Release II, Catalog No. 96011, Research Genetics, Inc.)

1. Description

Human BAC plasmid DNA pools are provided in 96-well microtiter plates: one "Superpool" plate, four plate pool plates, and 24 row/column pool plates. Each superpool consists of plasmid DNA prepared from clones arrayed into eight 384-well microtiter plates. This arrangement allows one to locate a plate containing a clone with as few as 56 PCR reactions. An additional 40 PCR reactions of the row and column pools will result in the exact well location of a clone. The vector used to construct these clones is pBeloBAC11. The average insert size of each clone is 130 kb, with a range of 90-300 kb. There are 147,456 clones in Release II, corresponding to 4-5 human genome equivalents for the library.

The Arraying of the human BAC DNA pools (see **Figure 12**):

- Superpools:

  At the superpool level, each pool is comprised of all the clones from a block of 8 microtiter plates. Superpool #1 consists of plates 1-8; superpool #2 consists of plates 9-16; and so on. There are total 48 superpools.

- Plate pools

  A plate pool is simply a plasmid preparation of all the clones from a single plate.

There are 384 plate pools arrayed into four 96-well microtiter plates. Each column of the 96-well plates represents the eight plate pools contained in one superpool. For example, column one of plate pool plate #1 holds the plate pools 1-8 while column two of that same plate pool plate holds the plate pools 9-16. There are a total of 384 plate pools and four plate pool plates.

* Row and column pools

Row and column pools are made by combining all of the like rows and columns, respectively, of each set of eight plates which compromise each superpool. They are arrayed in the 96-well plate such that two sets of row and column pools, each clearly labeled, are present. Carefully check the label on the top of the microtiter plate to be certain you will be testing the row and columns for the appropriate series of plate pools. The correct rows and columns are located in the wells directly under the label. There are 24 plates which contain row and column pools.

2. Procedure

A. Begin by testing each of the DNA pools in the plate labeled "Superpools". There will be 48 PCR reactions in this step (see following worksheet). These pools represent 4-5X coverage of the human genome so one should get around 4 positive "hits" (I found an average of about 4.5 hits per screening).

B. Use the plate pool screening guide to determine which plate(s) to test based on the results of the superpool screening.

C. Set up a PCR for each plate represented in the positive superpools. The row and

column pools associated with the plate pool may be set up at the same time (see following worksheet). The appropriate set of row and column pools may be located by finding the row/column pool label that lists the set of eight plates that are being screened at the plate pool level.

PCR HEADING: Amplify BAC Superpools with marker(s): _____ **DATE:** / / **PAGE #:** _____

Condition: _____ : _____ - _____ °C TD w/30 sec ext @ 72 °C (35X, 37X, 40X) + 94 °C 4' start

DNA Conc (µg/ml): ? **DNA Vol/Rxn (µl):** 0.56 **Total Rxn Vol (µl):** 14

PCR REACTION PARAMETERS:

LOCUS/GROUPING: _____

	µl/rxn	56
sterile D.I. H_2O:	8.4	470.40
1.25 mM dNTPs:	2.24	125.44
10X PCR buffer:	1.4	78.40
25 mM MgCl:	0.84	47.04
20 mM forward primer A:	0.175	9.80
20 mM reverse primer A:	0.175	9.80
25 mM spermidine:	0.14	7.84
Taq polymerase (5U/µl):	0.07	3.92
100% DMSO:		
TOTAL PREMIX VOLUME (µl):	13.44	752.64

_____ **Approx. Size (bp)**

2 **µl 10x Stop Buffer (xcff only)**

4 **µl 10 mg/ml EtBr in 75 ml gel**

3 **µl 100 bp ladder (50ng/µl)**

Time	
From:	To:

_____ **Start setup**

_____ **PCR**

_____ **Gel (100V, max Amp)**

_____ **Hits**

_____ **? hits**

_____ **Entered in BAC db**

Run products out on 2% 0.5X TBE agarose gels.

PCR rack set-up (BAC Superpool plate):

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	genomic					
B	2	10	18	26	34	42	genomic					
C	3	11	19	27	35	43	H_2O					
D	4	12	20	28	36	44						
E	5	13	21	29	37	45						
F	6	14	22	30	38	46						
G	7	15	23	31	39	47						
H	8	16	24	32	40	48						
	top		middle		bottom							

Gel top (L100 = 100 bp ladder)

top L-01-09-02-10-03-11-04-12-05-13-06-14-07-15-08-16-genomic

middle L-17-25-18-26-19-27-20-28-21-29-22-30-23-31-24-32-H2O

bottom L-33-41-34-42-35-43-36-44-37-45-38-46-39-47-40-48-genomic

Worksheet 2. BAC Screening Superpool PCR Setup.

PCR HEADING: Amplify BAC Plate/Row/Column pools with marker: **DATE:** / / **PAGE #:**

Condition: : - °C TD w/30 sec ext @ 72 °C (35X, 37X, 40X) + 94 °C 4' start

DNA Conc (µg/ml): ? **DNA Vol/Rxn (µl):** 1.12 **Total Rxn Vol (µl):** 14

PCR REACTION PARAMETERS:

LOCUS/GROUPING:

	µl/rxn	56
sterile D.I. H_2O:	7.84	439.04
1.25 mM dNTPs:	2.24	125.44
10X PCR buffer:	1.4	78.40
25 mM MgCl:	0.84	47.04
20 mM forward primer:	0.175	9.80
20 mM reverse primer:	0.175	9.80
25 mM spermidine:	0.14	7.84
Taq polymerase (5U/µl):	0.07	3.92
100% DMSO:		
TOTAL PREMIX VOLUME (µl):	12.88	721.28

_____ **Approx. Size (bp)**

2 **µl 10x Stop Buffer (xcff only)**

4 **µl 10 mg/ml EtBr in 75 ml gel**

3 **µl 100 bp ladder (50ng/µl)**

Time	
From:	To:

_____ **Start setup**

_____ **PCR**

_____ **Gel (100V, max Amp)**

_____ **Hits**

_____ **? hits**

_____ **Entered in BAC db**

Run products out on 2% 0.5X TBE agarose gels.

PCR rack set-up (BAC Plate Pools & Row/Column Pools):

	1	2	3	4	5	6	7	8	9	10	11	12
A	A	I	24	16	8	4	genomic					
B	B	J	23	15	7	3	genomic					
C	C	K	22	14	6	2	H_2O					
D	D	L	21	13	5	1						
E	E	M	20	12	P1-	P5-						
F	F	N	19	11	P2-	P6-						
G	G	O	18	10	P3-	P7-						
H	H	P	17	9	P4-	P8-						

top row/col. pool plate____ middle cols.____ bottom plate pool #____ column #____

BAC Address (P-R-C)_____

Gel top (L100 = 100 bp ladder)

top L100-A-I-B-J-C-K-D-L-E-M-F-N-G-O-H-P-genomic

middle L100-24-16-23-15-22-14-21-13-20-12-19-11-18-10-17-9-H2O

bottom L100-8-4-7-3-6-2-5-1-P1-P5-P2-P6-P3-P7-P4-P8-genomic

Worksheet 3. BAC Screening Plate/Row/Column Pool PCR Setup.

3. PCR Conditions

A. PCR conditions will vary according to primer length and base composition, reaction volume, and the type of temperature cycler used. Research Genetics, Inc. recommends using one 1 µl of plasmid superpool DNA or 2 µl of plasmid plate pool and row/column pool DNA in a final PCR reaction volume of 25 µl.

- I found that reducing the final PCR reaction volume to 14 µl (and proportionally the plasmid DNAs) was sufficient. This volume was still enough to detect positive hits on agarose. Reducing the amount of DNA per reaction increased the total number of times I could screen the pools.

- PCR primer concentrations, number of PCR cycles and annealing temperatures vary for each marker. It is best to work these out with test reactions first. I found that using 0.175 µl for each primer (20 µM) in a 14 µl reaction was typically sufficient. The best combination for ease in scoring the gel is a low primer concentration and a low number of PCR cycles.

- Include a 4-5 minute hold at 94°C as the first step of the PCR reaction to help maximize denaturation of DNA.

B. An automated multi-channel pipette (0.2-10 µl, Rainin Instrument Company) can be used to load plasmid DNA into the PCR reaction tubes to minimize the amount of time the plates are left unfrozen. When measuring small quantities it is important to ensure that all pipette tips are snugly in place. It is also much easier

to pipette small quantities of DNA into tubes already containing the PCR premix.

C. When analyzing the PCR reactions on a gel (generally 0.5X TBE 2% agarose), it is necessary to have included both genomic and water controls in your PCR set up. A DNA ladder should also be run alongside for size comparison. The automated multi-channel pipette can be used to add samples into the gel in order to minimize loading time.

- By arranging the DNA samples in a staggered format (as seen in the worksheets), an eight-channel pipette can be used to transfer the PCR products to an agarose gel (with six rows of 18 wells each) such that every other well is loaded for each column on the PCR plate. Each row on the gel will comprise of two adjacent columns on the PCR plate. With this method, the gels can be quickly loaded and the pools remain in numerical order. This also leaves room for a size control on both sides of each row. All reactions can be run at the same time on just one gel (just keep a careful watch to make sure one row does not run into another).

Human BAC DNA Superpool Release II Screening Guide

Use this guide when you have completed the 48 PCR reactions of the plates labeled "Superpools 1-48" to determine which plate pools must be tested.

Superpool	Superpool address	Plate Pool plate	Plate Pool wells	Correspond to plates...	Row/Column Pool plate	Row/Column Pool plate wells
1	1A	1	1A - 1H	1-8	1	1A - 5H
2	1B	1	2A - 2H	9-16	1	8A - 12H
3	1C	1	3A - 3H	17-24	2	1A - 5H
4	1D	1	4A - 4H	25-32	2	8A - 12H
5	1E	1	5A - 5H	33-40	3	1A - 5H
6	1F	1	6A - 6H	41-48	3	8A - 12H
7	1G	1	7A - 7H	49-56	4	1A - 5H
8	1H	1	8A - 8H	57-64	4	8A - 12H
9	2A	1	9A - 9H	65-72	5	1A - 5H
10	2B	1	10A - 10H	73-80	5	8A - 12H
11	2C	1	11A - 11H	81-88	6	1A - 5H
12	2D	1	12A - 12H	89-96	6	8A - 12H
13	2E	2	1A - 1H	97-104	7	1A - 5H
14	2F	2	2A - 2H	105-112	7	8A - 12H
15	2G	2	3A - 3H	113-120	8	1A - 5H
16	2H	2	4A - 4H	121-128	8	8A - 12H
17	3A	2	5A - 5H	129-136	9	1A - 5H
18	3B	2	6A - 6H	137-144	9	8A - 12H
19	3C	2	7A - 7H	145-152	10	1A - 5H
20	3D	2	8A - 8H	153-160	10	8A - 12H
21	3E	2	9A - 9H	161-168	11	1A - 5H
22	3F	2	10A - 10H	169-176	11	8A - 12H
23	3G	2	11A - 11H	177-184	12	1A - 5H
24	3H	2	12A - 12H	185-192	12	8A - 12H
25	4A	3	1A - 1H	193-200	13	1A - 5H
26	4B	3	2A - 2H	201-208	13	8A - 12H
27	4C	3	3A - 3H	209-216	14	1A - 5H
28	4D	3	4A - 4H	217-224	14	8A - 12H
29	4E	3	5A - 5H	225-232	15	1A - 5H
30	4F	3	6A - 6H	233-240	15	8A - 12H
31	4G	3	7A - 7H	241-248	16	1A - 5H
32	4H	3	8A - 8H	249-256	16	8A - 12H
33	5A	3	9A - 9H	257-264	17	1A - 5H
34	5B	3	10A - 10H	265-272	17	8A - 12H
35	5C	3	11A - 11H	273-280	18	1A - 5H
36	5D	3	12A - 12H	281-288	18	8A - 12H
37	5E	4	1A - 1H	289-296	19	1A - 5H
38	5F	4	2A - 2H	297-304	19	8A - 12H
39	5G	4	3A - 3H	305-312	20	1A - 5H
40	5H	4	4A - 4H	313-320	20	8A - 12H
41	6A	4	5A - 5H	321-328	21	1A - 5H
42	6B	4	6A - 6H	329-336	21	8A - 12H
43	6C	4	7A - 7H	337-344	22	1A - 5H
44	6D	4	8A - 8H	345-352	22	8A - 12H
45	6E	4	9A - 9H	353-360	23	1A - 5H
46	6F	4	10A - 10H	361-368	23	8A - 12H
47	6G	4	11A - 11H	369-376	24	1A - 5H
48	6H	4	12A - 12H	377-384	24	8A - 12H

Table 9. Human BAC DNA Superpool Release II Screening Guide.

Superpools Plate (1) - 48 reactions

(8 plates/pool x 384 clones/plate = 3072 clones/pool)
(3072 clones/pool x 48 pools = 147,456 total clones)

PCR - 0.56 µl DNA / 14 ml reaction

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41						
B	2	10	18	26	34	42						
C	3	11	19	27	35	43						
D	4	12	20	28	36	44	←		empty		→	
E	5	13	21	29	37	45						
F	6	14	22	30	38	46						
G	7	15	23	31	39	47						
H	8	16	24	32	40	48						

Row/Column Pools Plates (24) - 40 reactions

Row Pools: 8 plates/pool x 1 row/plate x 24 clones/row = 192 clones/pool
(192 clones/pool x 24 plates x 2 sets/plate x 16 pools/set = 147,456 total clones)

Column Pools: 8 plates/pool x 1 column/plate x 16 clones/column = 128 clones/pool
(128 clones/pool x 24 plates x 2 sets/plate x 24 pools/set = 147,456 total clones)

PCR - 1.12 µl DNA / 14 ml reaction

	Rows		**Columns**					**Rows**		**Columns**		
	1	2	3	4	5	6	7	8	9	10	11	12
A	A	I	24	16	8			A	I	24	16	8
B	B	J	23	15	7			B	J	23	15	7
C	C	K	22	14	6			C	K	22	14	6
D	D	L	21	13	5	empty		D	L	21	13	5
E	E	M	20	12	4			E	M	20	12	4
F	F	N	19	11	3			F	N	19	11	3
G	G	O	18	10	2			G	O	18	10	2
H	H	P	17	9	1			H	P	17	9	1

Figure 12. BAC Pool Arrays.

Plate Pools Plates (4) - 8 reactions

(384 clones/pool)
(384 clones/pool x 4 plates x 96 pools/plate = 147,456 total clones)

PCR - 1.12 µl DNA / 14 ml reaction

		1	2	3	4	5	6	7	8	9	10	11	12
	A	1	9	17	25	33	41	49	57	65	73	81	89
	B	2	10	18	26	34	X	50	58	66	74	82	90
	C	3	11	19	27	35	43	51	59	67	75	83	91
1	D	4	12	20	28	36	44	52	60	68	76	84	92
	E	5	13	21	X	37	45	53	61	69	77	85	93
	F	6	14	22	30	38	46	54	62	70	78	86	94
	G	7	15	23	31	39	47	55	63	71	79	87	95
	H	8	16	24	32	40	48	56	64	72	80	88	96

		1	2	3	4	5	6	7	8	9	10	11	12
	A	97	105	113	121	129	137	145	153	161	169	177	185
	B	98	106	114	122	130	138	146	154	162	170	178	186
	C	99	107	115	123	131	139	147	155	163	171	179	187
2	D	100	108	116	124	132	140	148	156	164	172	180	188
	E	101	109	117	125	133	141	149	157	165	173	181	189
	F	102	110	118	126	134	142	150	158	166	174	182	190
	G	103	111	119	127	135	143	151	159	167	175	183	191
	H	104	112	120	128	136	144	152	160	168	176	184	192

		1	2	3	4	5	6	7	8	9	10	11	12
	A	193	201	209	217	225	233	241	249	257	265	273	281
	B	194	202	210	218	226	234	242	250	258	266	274	282
	C	195	203	211	219	227	235	243	251	259	267	275	283
3	D	196	204	212	220	228	236	244	252	260	268	276	284
	E	197	205	213	221	229	237	245	253	261	269	277	285
	F	198	206	214	222	230	238	246	254	262	270	278	286
	G	199	207	215	223	231	239	247	255	263	271	279	287
	H	200	208	216	224	232	240	248	256	264	272	280	288

		1	2	3	4	5	6	7	8	9	10	11	12
	A	289	297	305	313	321	329	337	345	353	361	369	377
	B	290	298	306	314	322	330	338	346	354	362	370	378
	C	291	299	307	315	323	331	339	347	355	363	371	379
4	D	292	300	308	316	324	332	340	348	356	364	372	380
	E	293	301	309	317	325	333	341	349	357	365	373	381
	F	294	302	310	318	326	334	342	350	358	366	374	382
	G	295	303	311	319	327	335	343	351	359	367	375	383
	H	296	304	312	320	328	336	344	352	360	368	376	384

Figure 12. BAC Pool Arrays (continued).

Superpools Plate - 48 reactions

(Here we have 4 positive hits, we'll follow through with superpool #3)

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41						
B	2	10	18	26	34	42						
C	3	11	19	27	35	43						
D	4	12	20	28	36	44	←		empty			→
E	5	13	21	29	37	45						
F	6	14	22	30	38	46						
G	7	15	23	31	39	47						
H	8	16	24	32	40	48						

Plate Pools Plate #1 - 8 reactions

(Superpool #3 corresponds to Plate Pool Plate #1, column 3. Here we get 1 positive hit, plate #18)

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	95
H	8	16	24	32	40	48	56	64	72	80	88	96

Row/Column Pools Plate #2 - 40 reactions

(Superpool #3 corresponds to Row/Column Pool Plate #2, columns 1-5. Here we get 2 positive hits, row K and column #8)

	1	2	3	4	5	6	7	8	9	10	11	12
A	A	I	24	16	8			A	I	24	16	8
B	B	J	23	15	7			B	J	23	15	7
C	C	K	22	14	6			C	K	22	14	6
D	D	L	21	13	5	empty		D	L	21	13	5
E	E	M	20	12	4			E	M	20	12	4
F	F	N	19	11	3			F	N	19	11	3
G	G	O	18	10	2			G	O	18	10	2
H	H	P	17	9	1			H	P	17	9	1

Our BAC address is: 18 - K - 8 (Plate 18 - Row K - Column 8)

Figure 13. BAC Screening Example.

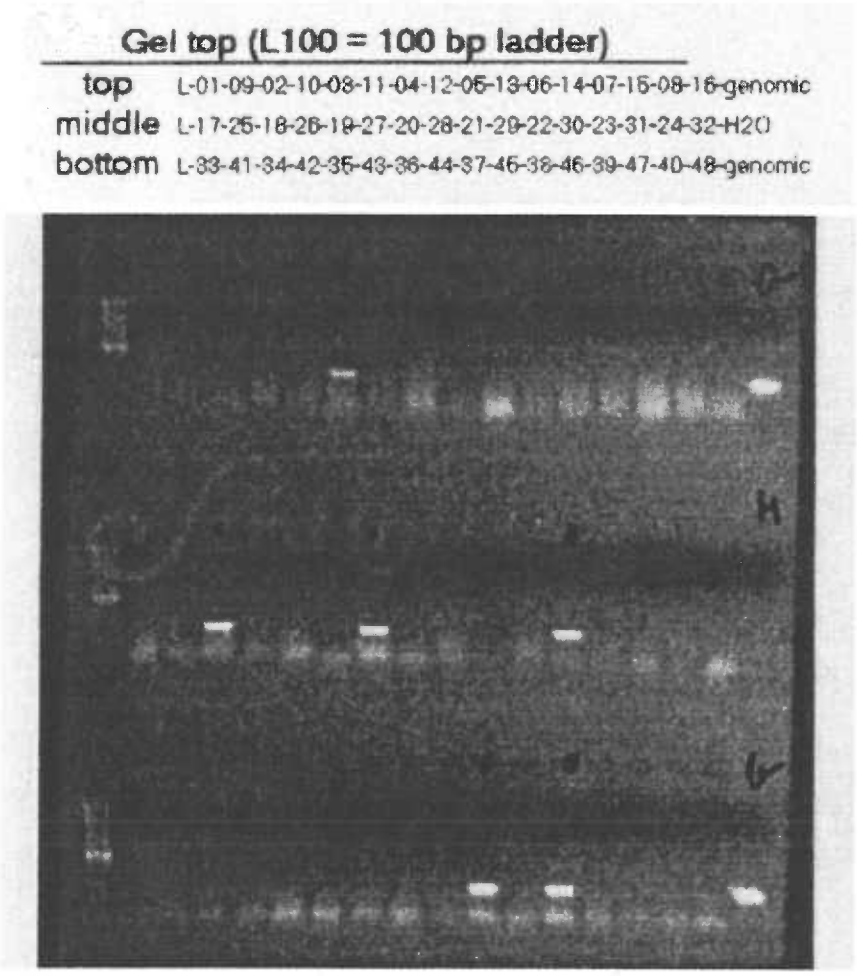


Figure 14. An ethidium bromide stained agarose gel of PCR products generated from the Research Genetics, Inc. human BAC superpool plate using marker 142L14-75. This STS marker was obtained by end sequencing BAC 142-L-14 (Research Genetics, Inc.) with the pUC/M13 Forward (-47) primer. A 100 bp ladder was run in the first well of each row. The last sample in each row is genomic control, H2O control, genomic control, respectively.

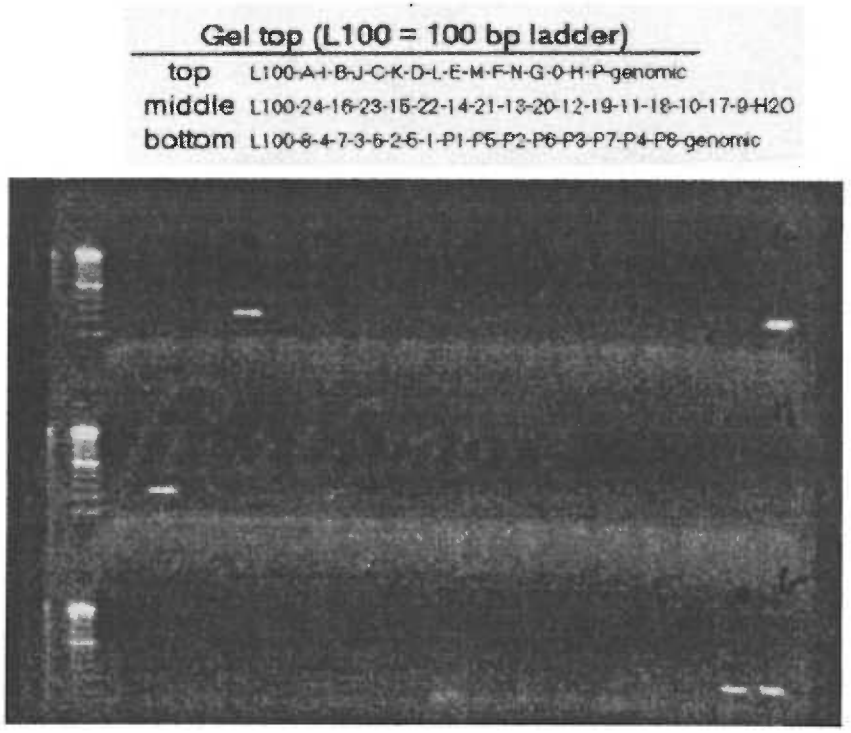


Figure 15. An ethidium bromide stained agarose gel of PCR products generated from the Research Genetics, Inc. human BAC plate and row/column plate pools using marker D20S906. This set of pools was used in correspondence to a positive hit in superpool well #44 in the first round of screening. The hit in top row, lane 5 corresponds to row J; the hit in middle row, lane 3 corresponds to column 16; and the hit in bottom row, lane 17 corresponds to plate 352. The BAC address is 352-J-16. A 100 bp ladder was run in the first well of each row. The last sample in each row is genomic control, H2O control, genomic control, respectively.

BAC DNA EXTRACTION AND END SEQUENCING

1. Plating Out BAC clone

A. Streak a Luria-Bertani (LB) agar plate containing 12.5 μ g/ml chloramphenicol with a BAC clone (Research Genetics, Inc.) from an agar stab or microtiter plate.

B. Grow overnight in 37° C incubator (Lab-Line Instruments, Inc., Melrose Park IL, USA; VWR Scientific, West Chester PA, USA).

2. Verification of BAC clone

A. Choose five colonies per BAC clone for whole cell PCR. Select colonies with sterile pipette tip.

B. Soak pipette tip in 4 μl sterile H2O in a 0.2 ml thin-walled PCR microtube (Island Scientific, Bainbridge Island WA, USA; Perkin-Elmer Cetus Corp.).

C. Run PCR using BAC specific primers at their pre-determined PCR profile.

D. Add 10X stop buffer and run reactions on a 1-2% agarose gel containing 0.5 μg/ml ethidium bromide for 30 minutes at 100V.

3. BAC DNA Isolation

A. Grow 5 ml starter culture (LB broth with 12.5 μ g/ml chloramphenicol) with a positive colony chosen from previous screening overnight at 37° C with shaking motion (~200 rpm).

B. Inoculate 500 ml LB broth containing 12.5 μ g/ml chloramphenicol with 2 ml of starter culture.

C. Grow overnight at 37° C with shaking motion (~200 rpm).

D. Isolate per manufacturer's protocol (QIAGEN Maxi Plasmid Purification, QIAGEN Plasmid Handbook, February 1995) briefly outlined below.

- Resuspend the bacterial pellet in 10 ml of Buffer P1.
- Add 10 ml of Buffer P2, mix gently, and incubate at room temperature for 5 minutes.
- Add 10 ml of chilled Buffer P3, mix immediately but gently, and incubate on ice for 15-20 minutes.
- Centrifuge (5415, Brinkmann Instruments Inc., Westbury NY, USA) at ≥ 20,000 x g for 30 minutes at 4° C. Remove supernatant promptly.
- Centrifuge at ≥ 20,000 x g for 15 minutes at 4° C. Remove supernatant promptly. (Alternatively, the supernatant can be filtered over a prewetted, folded filter.)
- Equilibrate a QIAGEN-tip 500 by applying 10 ml Buffer QBT, and allow the column to empty by gravity flow.
- Apply the supernatant from step 5) to the QIAGEN-tip and allow it to enter the resin by gravity flow.
- Wash the QIAGEN-tip with 2 x 30 ml Buffer QC.

- Elute DNA with 15 ml Buffer QF.
- Precipitate DNA with 0.7 volumes of room-temperature isopropanol. Centrifuge immediately at ≥ 15,000 x g for 30 minutes at 4° C, and carefully remove the supernatant.
- Wash DNA with 5 ml of 70% ethanol, air-dry for 5 minutes, and redissolve in a suitable volume of buffer.
- Resuspend DNA in 1 ml H2O.
- Phenol-chloroform extract twice (24:1 chloroform:isoamyl alcohol) and chloroform:isoamyl alcohol extract once.
- Precipitate 1-2 hours at -20° C in 1/10 volume 3M Na-Acetate and 2X volume 95% ethanol.
- Spin 15 minutes at 4° C at 14,000 rpm.
- Pour off supernatant carefully; wash quickly with 500 µl 70% ethanol and let dry.
- Resuspend in H2O and quantitate.

4. **BAC End Sequencing** (OHSU Core Sequencing Facility, Portland OR, USA)

A. Using automated sequencing reaction with Dye Terminator Mix/premix.

B. Mix 8 µl premix, 0.5 µ g BAC DNA, 3.2 pmols of primer, and enough H2O to bring total volume to 20 µl in a small, thin-walled tube. Primers used are

either pUC/M13 Forward (-47) 5'-

d(CGCCAGGGTTTTCCCAGTCACGAC)-3' or pUC/M13 Reverse 5'-

d(TCACACAGGAAACAGCTATGAC)-3'.

C. Download sequence tracing and text files. Analyze.

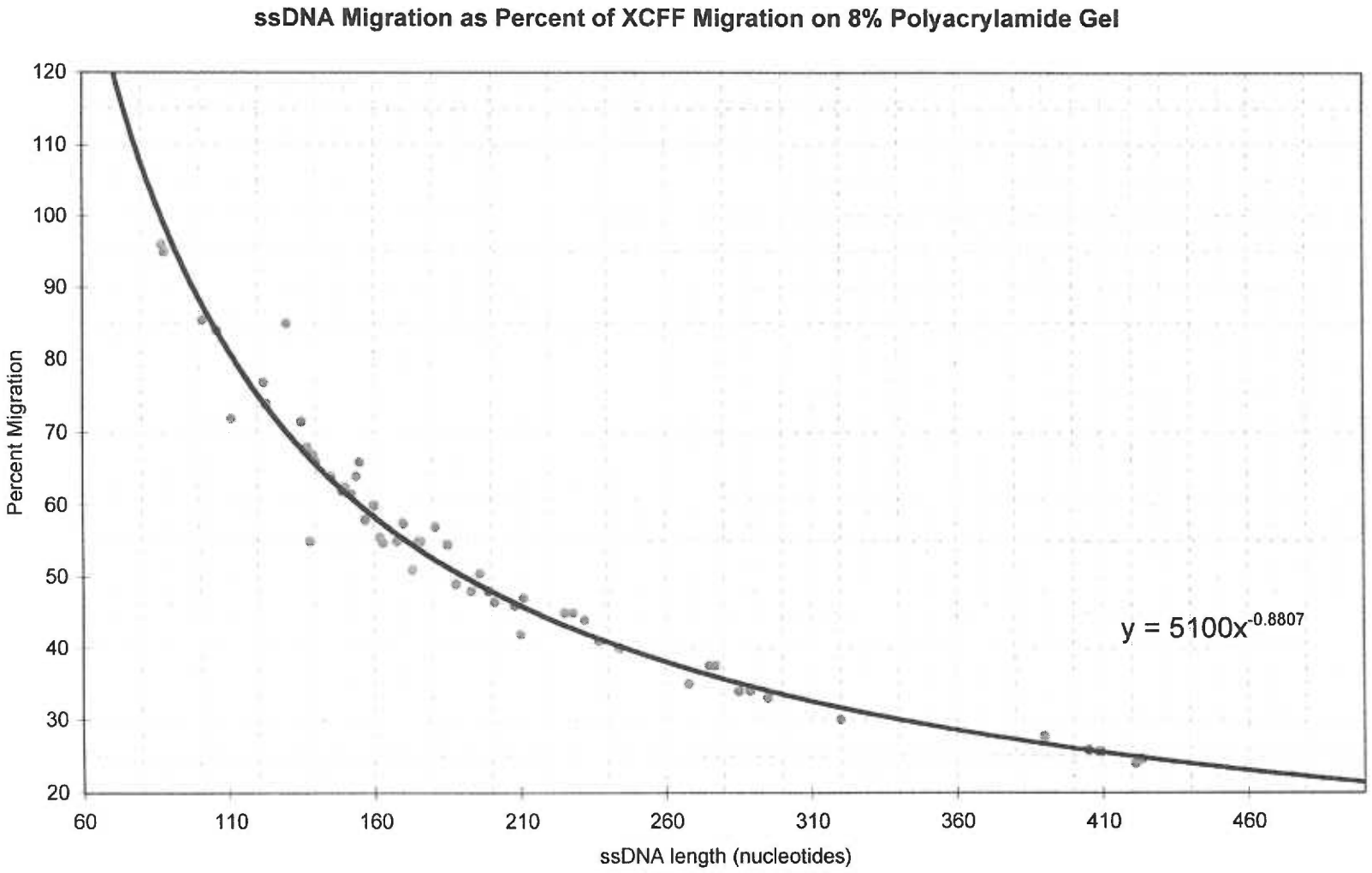


Figure 16. Migration Chart.

OTHER PAPERS

Clouston Syndrome (Hidrotic Ectodermal Dysplasia) Is Not Linked to Keratin Gene Clusters on Chromosomes 12 and 17

Susan J. Hayflick,* Todd Taylor,* Wendy McKinnon,† Alan E. Guttmacher,† Mike Litt,* and Jonathan Zonana*

*Department of Molecular and Medical Genetics, Oregon Health Sciences University, Portland, U.S.A.; and †Department of Pediatrics, University of Vermont, Burlington, U.S.A.

Clouston syndrome is an hidrotic form of ectodermal dysplasia, inherited as an autosomal dominant trait with high penetrance. The main features of the disorder are alopecia, severe dystrophy of the nails, and palmoplantar hyperkeratosis. A molecular abnormality of keratin has long been hypothesized to be the basic defect in this disorder. We have performed linkage analyses between the disorder and markers close to the keratin gene clusters on chromosomes 12 and 17 and have excluded linkage to these candidate regions in three apparently unrelated families. In addition, linkage has been excluded to four other candidate regions including 1q21, 17q23-qter, 18q21, and 20q12. These data indicate that Clouston syndrome is not due to a defect in keratin or in a subset of keratin-associated proteins. ***Key words: hyperkeratosis/gene mapping/linkage/microsatellite markers. J Invest Dermatol 107:11–14, 1996***

Clouston syndrome (Mendelian Influence in Man (MIM) 129500) is an autosomal dominant ectodermal dysplasia characterized by the triad of nail dystrophy, alopecia, and hyperkeratosis of the palms and soles (Clouston, 1929, 1939). In contrast to the X-linked form of ectodermal dysplasia, teeth, facial appearance, and sweating are normal. The pathogenesis is unknown. In 1965, Scriver *et al* suggested that a molecular defect in keratin was the cause; this hypothesis was supported by the biochemical identification of depleted hair matrix proteins accompanied by disruption of or failure to form disulfide bonds in the remaining keratin molecules of hair (Gold and Scriver, 1971, 1972). These findings implicated a gene coding for a component of the high sulfur fraction of matrix α-keratin. Further support for a possible role of keratin came from ultrastructural studies of hair from affected individuals (Escobar *et al*, 1983), which demonstrated disorganization of fibrils with loss of cuticular cortex. In addition, other disorders with palmoplantar hyperkeratosis are caused by mutations in keratin genes, including epidermolytic palmoplantar keratoderma (Reis *et al*, 1992), nonepidermolytic palmoplantar keratoderma (Kimonis *et al*, 1994), and pachonychia congenita (McLean *et al*, 1995), and may even be allelic to Clouston syndrome. Based on these findings, candidate genes for Clouston syndrome include the keratin gene family.

Keratins are intermediate filaments that form a complex and elaborate cytoskeleton essential to the structural integrity of individual epithelial cells as well as whole tissues. Keratin intermediate filaments are heterodimers formed of one type I, or acidic, keratin protein chain and one type II, or basic, chain. The type I keratins are encoded by a gene family located on chromosome 17q21.1-q21.2 (Romano *et al*, 1988; Rosenberg *et al*, 1988; Savtchenko *et al*, 1990), and the type II keratin gene family is localized to chromosome 12q11-q13 (Romano *et al*, 1988; Yoon *et al*, 1994).

In order to determine whether defects in keratin proteins are associated with Clouston syndrome, we analyzed the cosegregation of highly polymorphic microsatellite markers near the keratin gene clusters with the Clouston syndrome phenotype in three unrelated families.

Four additional candidate regions (1q21, 17q23-qter, 18q21, and 20q12) were studied, including two that are implicated in disorders of palmoplantar hyperkeratosis and two that contain genes for keratin-associated proteins. Desmogleins, members of the cadherin superfamily, are adhesive proteins of the desmosome cell junctions in epithelial cells and form a 3-dimensional lattice with keratin filaments. Three desmoglein genes map to chromosome 18q12.1 (Simrak *et al*, 1994; Wang *et al*, 1994a), the region to which striate palmoplantar keratoderma has been linked (Hennies *et al*, 1995a). Keratoderma with tylosis and esophageal cancer maps to 17q23-qter, separate from the keratin gene cluster (Hennies *et al*, 1995b). Six genes expressed in terminally differentiating epidermis (trichohyalin, profilaggrin, involucrin, loricrin, small proline-rich protein, and calcyclin) map to a 2-megabase region on chromosome 1q21 (Fietz *et al*, 1992; Lee *et al*, 1993). Transglutaminases catalyze the formation of lysine isodipeptide crosslinks in proteins. TGM3 is found in terminally differentiating epidermal and hair kertinocytes and maps to chromosome 20q12 (Wang *et al*, 1995b). Markers linked to these candidate regions were studied for cosegregation with the Clouston syndrome phenotype.

MATERIALS AND METHODS

Families were ascertained through clinical genetics centers and were examined by a clinical geneticist (**Fig 1**). Families were identified with Clouston syndrome when 1) affected members showed at least two of the

Manuscript received November 30, 1995; revised February 29, 1996; accepted for publication March 5, 1996.

Reprint requests to: Dr. Susan J. Hayflick, Molecular and Medical Genetics, mail code L103, 3181 SW Sam Jackson Park Rd., Portland, OR 97201-3098.

Abbreviation: PCR, polymerase chain reaction.

0022-202X/96/$10.50 ·

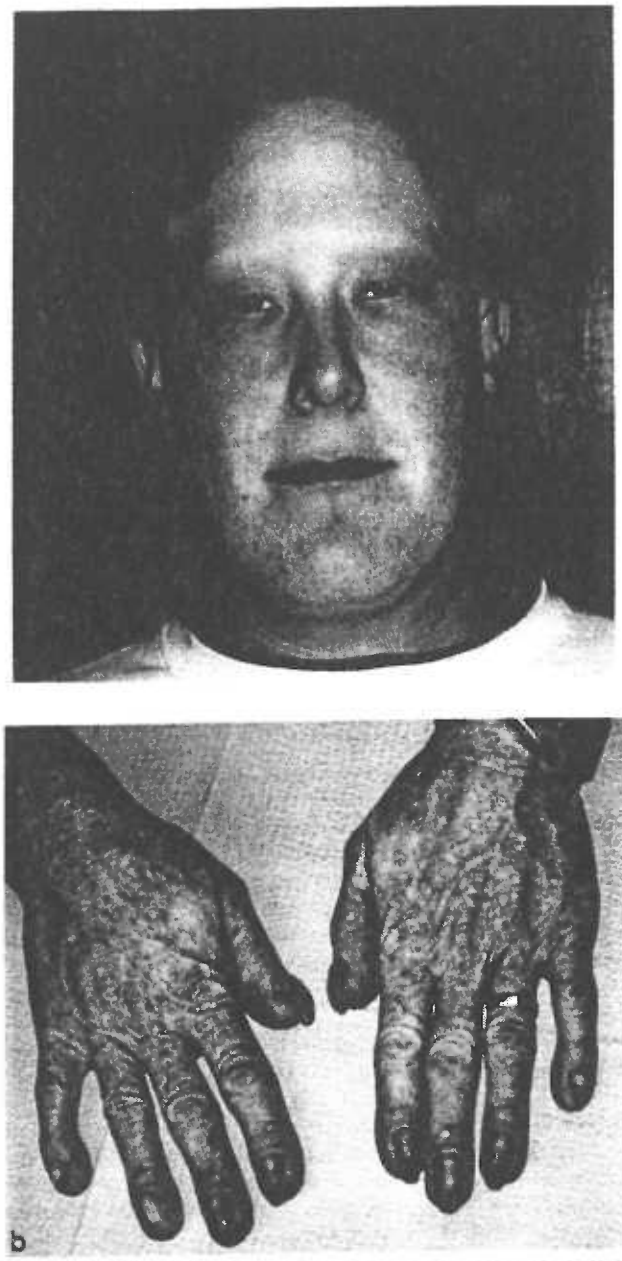


Figure 1. *a*, **absent scalp hair, facial hair and eyebrows in individual CS1 III-406.** *b*, **nail dystrophy in individual CS1 II-301.**

following features: sparse to absent scalp hair, dystrophic thickened nails, and thickening of the skin on the palms or plantar surfaces, 2) the disease followed a pattern of autosomal dominant inheritance, and 3) hypohidrosis or anhidrosis was absent.

Informed consent to participate in these studies was obtained from all participants. Fifteen milliliters of whole blood was obtained, and DNA was isolated by standard procedures (Miller *et al*, 1988). DNA was collected from a total of 35 individuals in three unrelated kindreds with a total of 18 affected individuals. The CS1 family is of Scottish and Irish descent (see **Fig 2** for pedigrees, maximum possible lod = 3.61, $\theta = 0$). The CS2 (max possible lod = 1.20) and CS3 (max possible lod = 1.51) families are of French-Canadian descent and share a surname, though they are not known be related.

For linkage analysis, polymorphic markers linked to the keratin gene clusters and adjacent regions on chromosomes 12 and 17 and those linked to chromosomes 1q21, 17q23-qter, 18q21, and 20q12 were studied. All markers are polymerase chain reaction (PCR)-based short tandem repeat, microsatellite markers (Gyapay *et al*, 1994). Selected markers had an average heterozygosity of 76%.

PCR was performed in a total volume of 7 μl with 14 ng of genomic DNA, 3.5 pmol of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM

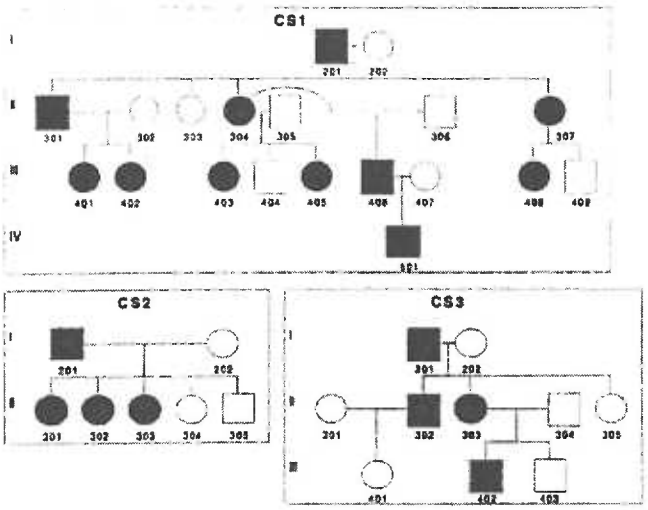


Figure 2. Pedigrees of the families studied.

KCl, 10 mM Tris-Cl, pH 8.5, and 0.175 U of *Taq* polymerase (Amplitaq). In a typical experiment, two-temperature "touchdown" PCR amplification was performed in a Perkin-Elmer/Cetus System 9600 thermocycler, using an initial annealing temperature of 65°C, which was decreased by one degree in each of the first 15 cycles, then maintained at 50°C for 20 more cycles. Each cycle consisted of a 15-s, 94°C denaturing step and a 30-s annealing/extension step. This was followed by a final 5-min extension step at 72°C, and the temperature was then reduced to 15°C until the reaction tubes were removed from the apparatus.

After PCR amplification, products were resolved on DNA sequencing gels containing 5.6 M urea and 32% formamide, transferred to nylon membranes by capillary blotting and revealed by probing the membranes with 5'-[^{32}P]-labeled $(CA)_{15}$ as previously described (Litt *et al*, 1993).

Linkage analysis was performed using the MLINK and LINKMAP options of the LINKAGE package (Lathrop *et al*, 1984; Terwilliger and Ott, 1994). Autosomal dominant inheritance with 95% penetrance was assumed using a population gene frequency of 10^{-5}. Because reliable population allele frequencies for most of the markers do not exist, they were assumed to be equal for lod score calculations. Genetic distances between the markers were based on published information (National Institutes of Health/CEPH Collaborative Mapping Group, 1992; Gyapay *et al*, 1994) or were calculated using genotypes in the CEPH database v7.0. These genetic distances agreed with those calculated using genotypes of our families. When pairwise analysis was uninformative, multipoint analysis was performed. Because of the possibility of locus heterogeneity, linkage analysis was performed independently on each family.

RESULTS

Table I shows combined lod scores obtained in the three families using the designated markers. Individual family data gave no evidence for linkage to the regions studied. These data exclude close linkage to markers in the regions spanning the keratin gene clusters and additional candidate regions on chromosomes 1q21, 17q23-qter, 18q21, and 20q12. Using the criterion that a lod score ≤ -2 excludes linkage together with the published linkage maps of the regions (Gyapay *et al*, 1994; Yoon *et al*, 1994), we can exclude each of the regions studied.

DISCUSSION

Based on biochemical and pathologic studies, keratin defects have been implicated in the etiology of Clouston syndrome. Therefore the genes encoding keratin proteins are candidate genes in this disease. Additional candidates include genes for other disorders of hyperkeratosis, with which Clouston syndrome might be allelic, and regions to which genes that encode keratin-associated proteins have been mapped. The objective of our study was to investigate co-segregation of Clouston syndrome with highly polymorphic markers tightly linked to these regions. Linkage was excluded to the keratin gene clusters on chromosomes 12 and 17 in three

Table 1. Summed Lod Scores for the CS Families at Markers Near the Keratin Genes and Other Candidate Loci[a]

		Lod score at theta =							
cM	Marker	0.00	0.01	0.05	0.10	0.20	0.30	0.40	Candidate Genes
1	D1S534	−6.69	−2.45	−1.26	−0.63	−0.05	0.05	0.15	Trichohyalin, profilaggrin, involucrin, loricrin
4	D1S442	−12.34	−4.45	−2.35	−1.27	−0.27	0.11	0.17	
	D1S305	−18.83	−6.94	−3.73	−2.30	−1.03	−0.48	−0.20	
<1	D12S368	−17.23	−6.16	−2.97	−1.57	−0.38	0.04	0.11	- KRT7
2	D12S262	−8.71	−1.15	0.23	0.70	0.88	0.67	0.32	- KRT1, KRT2, KRT5, KRT6a, KRT6b
3	D12S103	−13.88	−5.04	−2.66	−1.59	−0.64	−0.23	−0.05	- KRT4, KRT8, KRT18
	D12S90	−11.28	−2.79	−0.99	−0.19	0.39	0.45	0.29	
6	D17S783	−6.56	−2.60	−1.56	−1.01	−0.47	−0.21	−0.08	
	D17S798	−23.07	−8.04	−4.13	−2.38	−0.80	−0.12	0.11	
9	D17S250	−6.54	−2.23	−1.34	−0.90	−0.45	−0.22	−0.08	
	D17S800	−11.34	−3.71	−1.56	−0.70	−0.02	0.18	0.17	
3	KT10	−12.61	−6.06	−3.10	−1.85	−0.72	−0.22	−0.01	- KT10
	KRT9	−21.37	−8.69	−4.68	−2.83	−1.09	−0.30	0.02	- KRT9
3	D17S791	−29.29	−9.02	−4.88	−3.00	−1.25	−0.43	−0.06	
	D17S579	−23.86	−10.90	−5.77	−3.56	−1.51	−0.55	−0.10	
8	D17S806	−31.51	−10.61	−5.49	−3.30	−1.31	−0.41	−0.02	
8	D17S787	−15.74	−7.60	−4.30	−2.75	−1.24	−0.50	−0.12	
9	D17S808	−15.08	−2.15	−0.11	0.62	1.00	0.88	0.51	
4	D17S795	−18.11	−4.66	−2.03	−1.04	−0.27	−0.02	0.03	
11	D17S949	−21.78	−4.99	−1.86	−0.59	0.35	0.54	0.38	- TOCG
2	D17S785	−11.81	−1.90	−0.47	0.07	0.39	0.35	0.19	
11	D17S937	−6.79	−1.59	−0.72	−0.32	0.00	0.09	0.08	
	D17S784	−6.62	−6.41	−4.71	−3.15	−1.60	−0.80	−0.32	
2	D18S56	−12.27	−6.05	−3.15	−1.91	−0.78	−0.26	−0.03	- Striate PPK
2	DD18S457	−5.32	−2.53	−1.20	−0.69	−0.27	−0.09	−0.01	
2	D18S456	−17.93	−9.43	−6.10	−4.07	−2.08	−1.01	−0.37	
	D18S57	−19.72	−9.91	−5.54	−3.57	−1.69	−0.74	−0.23	
12	D20S112	−5.87	−1.46	−0.50	−0.07	0.23	0.23	0.13	- TGM3
9	D20S106	−18.75	−8.02	−4.18	−2.53	−1.07	−0.44	−0.14	
	D20S107	−19.27	−6.04	−2.90	−1.55	−0.43	0.00	0.09	

[a] Genetic distances between adjacent markers are listed under "cM" (centiMorgans).

families. Furthermore, no evidence for linkage was found to markers near genes implicated in two other disorders of hyperkeratosis, striate palmoplantar keratoderma and keratoderma with tylosis and esophageal cancer, or in regions to which genes for keratin-associated proteins localize.

Clouston's original family with hidrotic ectodermal dysplasia was French-Canadian, as are the CS2 and CS3 families, though none of these families is known to be related to the others. The third family we studied is of Scottish and Irish descent. Therefore, we have excluded the keratin gene clusters in families from two distinct ethnic groups. As in all genetic disorders, the possibility of nonallelic genetic heterogeneity must be considered in Clouston syndrome.

Several additional keratin-associated matrix proteins have been described, and the genes for these were mapped (Fratini *et al*, 1993; Rogers and Powell, 1993; Chou *et al*, 1994). Though the original hypothesis of a keratin gene mutation causing Clouston syndrome is not supported by our data, other candidate genes await investigation.

We thank Dr. Virginia Sybert for informing us about the CS1 family. We are grateful to the families and the National Foundation for Ectodermal Dysplasia, whose enthusiastic participation made this project possible. This work was supported by Oregon Health Sciences University Foundation Grant MRF9319 (S.J.H.) and National Institutes of Health Grant R01-DE11311-5 (J.Z.).

REFERENCES

Chou CF, Riopel CL, Omary MB: Identification of a keratin-associated protein that localizes to a membrane compartment. *Biochem J* 298:457-463, 1994

Clouston HR: A hereditary ectodermal dystrophy. *Canad Med Assoc J* 21:18-31, 1929

Clouston HR: The major forms of hereditary ectodermal dysplasia. *Canad Med Assoc J* 40:1-7, 1939

Escobar V, Goldblatt LI, Bixler D, Weaver D: Clouston syndrome: an ultrastructural study. *Clin Genet* 24:140-146, 1983

Fietz MJ, Rogers GE, Eyre HJ, Baker E, Callen DF, Sutherland GR: Mapping of the trichohyalin gene: co-localization with the profilaggrin, involucrin, and loricrin genes. *J Invest Dermatol* 99:542-544, 1992

Fratini A, Powell BC, Rogers GE: Sequence, expression, and evolutionary conservation of a gene encoding a glycine/tyrosine-rich keratin-associated protein of hair. *J Biol Chem* 268(6):4511-4518, 1993

Gold RJM, Scriver CR: The characterization of hereditary abnormalities of keratin: Clouston's ectodermal dysplasia. *Birth Defects Orig Art Ser* 6:91-95, 1971

Gold RJM, Scriver CR: Properties of hair keratin in an autosomal dominant form of ectodermal dysplasia. *Am J Hum Genet* 24:549-561, 1972

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J: 1994 The 1993-94 Généthon human genetic linkage map. *Nat Genet* 7:246-339

Hennies H-C, Hagedorn M, Reis A: Palmoplantar keratoderma in association with carcinoma of the esophagus maps to chromosome 17q distal to the keratin gene cluster. *Genomics* 29:537-540, 1995a

Hennies H-C, Kuster W, Mischke D, Reis A: Localization of a locus for the striated form of palmoplantar keratoderma to chromosome 18q near the desmosomal cadherin gene cluster. *Hum Mol Genet* 4:1015-1020, 1995b

Kimonis V, DiGiovanna JJ, Yang J-M, Doyle SZ, Bale SJ, Compton JG: A mutation in the V1 end domain of keratin 1 in non-epidermolytic palmar-plantar keratoderma. *J Invest Dermatol* 103:764-769, 1994

Lathrop GM, Lalouel JM: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443-3446, 1984

Lee SC, Wang M, McBride OW, O'Keefe EJ, Kim IG, Steinert PM: Human trichohyalin gene is clustered with the genes for other epidermal structural proteins and calcium-binding proteins at chromosomal locus 1q21. *J Invest Dermatol* 100:65-68, 1993

Litt M, Hauge XY, Sharma V: Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. *BioTechniques* 15:280-284, 1993

McLean WHI, Rugg EL, Lunny DP, Morley SM, Lane EB, Swensson O, Dopping-Hepenstal PJC, Griffiths WAD, Eady RAJ, Higgins C, Navasaria HA, Leigh IM,

Strachan T, Kunkeler L, Munro CS: Keratin 16 and keratin 17 mutations cause pachonychia congenita. *Nature Genet* 9:273–278, 1995

Miller SA, Dykes DD, Polesky: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215, 1988

NIH/CEPH Collaborative Mapping Group: A comprehensive genetic linkage map of the human genome. *Science* 258:67–86, 1992

Rogers GE, Powell BC: Organization and expression of hair follicle genes. *J Invest Dermatol* 101:50S–55S, 1993

Reis A, Kuster W, Eckhardt R, Sperling K: Mapping of a gene for epidermolytic palmoplantar keratoderma to the region of the acidic deratin gene cluster at 17q12–q21. *Hum Genet* 90:113–116, 1992

Romano V, Bosco P, Rocchi M, Costa G, Leube RE, Franke WW, Romeo G: Chromosomal assignments of human type I and type II cytokeratin genes to different chromosomes. *Cytogenet Cell Genet* 48:148–151, 1988

Rosenberg M, RayChaudhury A, Shows TB, Le Beau MM, Fuchs E: A group of type I keratin genes on human chromosome 17: characterization and expression. *Mol Cell Biol* 8:722–736, 1988

Savtchenko ES, Tomic M, Ivker R, Blumenberg M: Three parallel linkage groups of human acidic keratin genes. *Genomics* 7:394–407, 1990

Scriver CR, Solomons CC, Davies E, Williams M, Bolton J: A molecular abnormality of keratin in ectodermal dysplasia. *J Pediat* 67:946, 1965

Simrak D, Cowley CM, Buxton RS, Arnemann J: Tandem arrangement of the closely linked desmoglein genes on human chromosome 18. *Genomics* 25:591–594, 1995

Terwilliger JD, Ott J: In: *Handbook of Human Genetic Linkage*. Johns Hopkins University Press, Baltimore, 1994

Wang Y, Amagai M, Minoshima S, Sakai K, Green KJ, Nishikawa T, Shimizu N: The human genes for desmogleins (DSG1 and DSG3) are located in a small region on chromosome 18q12. *Genomics* 20:492–495, 1994

Wang M, Kim IG, Steinert PM, McBride OW: Assignment of the human transglutaminase 2 (TGM2) and transglutaminase 3 (TGM3) genes to chromosome 20q11.2. *Genomics* 23:721–722, 1994

Yoon SJ, LeBlanc-Straceski J, Straceski J, Ward D, Krauter K, Kucherlapati R: Organization of the human keratin type II gene cluster at 12q13. *Genomics* 24:502–508, 1994

Confirmation of Linkage of Clouston Syndrome (Hidrotic Ectodermal Dysplasia) to 13q11-q12.1 with Evidence for Multiple Independent Mutations

Todd D. Taylor, Susan J. Hayflick, Wendy McKinnon,* Alan E. Guttmacher,* Alain Hovnanian,† Mike Litt, and Jonathan Zonana

Department of Molecular and Medical Genetics, Oregon Health Sciences University, Portland, Oregon, U.S.A.; *Department of Pediatrics, University of Vermont, Burlington, Vermont, U.S.A.; †The Wellcome Trust Center for Human Genetics, University of Oxford, Oxford, U.K.

Clouston syndrome (hidrotic ectodermal dysplasia) is an autosomal dominant disorder characterized by the triad of nail dystrophy, alopecia, and palmoplantar hyperkeratosis. Recently, linkage of a Clouston syndrome locus to chromosome 13q11-q12.1 was reported in eight families of French-Canadian descent. We have confirmed linkage to this region in four additional families: two of French-Canadian descent, one of Scottish–Irish descent, and one French family. Multipoint linkage analysis gave a lod score of 5.09 at marker D13S175. The two families of French-Canadian descent share haplotypes with those reported by Kibar *et al* (1996), indicating a common founder. The French and Scottish–Irish families do not demonstrate the common haplotype, indicating that the mutations in these populations are most likely of different origin. ***Key words: gene mapping/hyperkeratosis/microsatellite markers. J Invest Dermatol 111:83–85, 1998***

Clouston syndrome (MIM no. 129500) is an hidrotic form of ectodermal dysplasia, inherited as an autosomal dominant trait with high penetrance. Clinical findings include dystrophic nails, alopecia, and palmoplantar hyperkeratosis (Clouston, 1929, 1939). Recent reports indicate that a locus for Clouston syndrome maps to the pericentromeric region of chromosome 13q in eight French–Canadian families (Kibar *et al*, 1996). A second large Indian family also shows linkage to the same region (Radhakrishna *et al*, 1997).

In this study, we confirm linkage of the Clouston syndrome gene to the pericentromeric region of chromosome 13q in an ethnically diverse group of families. We demonstrate further evidence for a common haplotype among the French-Canadian families, which is not shared by families of either French or Scottish–Irish descent.

MATERIALS AND METHODS

Patients Families CS1 (Scottish–Irish), CS2, and CS3 (both of French-Canadian descent) were previously described by Hayflick *et al* (1996). Family CS4 is a French family affected through five generations. An affected cousin of the mother was reported by Giraud *et al* (1977). This patient showed classical signs of Clouston syndrome, including sparse hair, nail dystrophy, and hyperkeratosis. Blood samples were collected from a total of 38 individuals in four unrelated kindreds with a total of 21 affected individuals (nine males and 12 females). The maximum possible lod score for these families is 6.62 at $\theta = 0$. DNA was isolated by standard procedures (Miller *et al*, 1988).

Genotyping Polymorphic markers from chromosome 13q (D13S1316, D13S175, D13S141, D13S143, D13S115, D13S1236, D13S1243, and D13S221) were studied. All markers typed were PCR-based short tandem repeat, microsatellite markers (Hudson *et al*, 1992; Petrukhin *et al*, 1993; Gyapay *et al*, 1994; Dib *et al*, 1996). Analysis of microsatellite markers was performed as previously described (Litt *et al*, 1993).

Linkage analysis Linkage analysis was performed using the LINKAGE 5.1/FASTLINK 3.0P packages (Lathrop *et al*, 1984; Cottingham *et al*, 1993; Schäffer *et al*, 1994; Schäffer, 1996). Clouston syndrome was modeled as a rare, autosomal dominant disorder with complete penetrance and an estimated gene frequency of 10^{-5}. Marker allele frequencies were derived from the genotypes of 13 unrelated spouses participating in this study and from the Center d'Etude du Polymorphisme Humain database. Marker order is based on a yeast artificial chromosome contig and expressed sequence tag map of the pericentromeric region of chromosome 13 (Guilford *et al*, 1995) and on the Généthon map.

RESULTS AND DISCUSSION

For marker D13S1236 the combined two-point lod score was 4.82 at $\theta = 0$ (**Table I**). Three other markers, D13S1316, D13S175, and D13S141, also showed no recombination with Clouston syndrome at $\theta = 0$ (with respective Z_{max} of 3.61, 2.53, and 2.80). Lod scores for markers D13S143 and D13S115 are $-\infty$ at $\theta = 0$ because there are obligate recombinants in families CS2 and CS3, respectively. The lod score for marker D13S1236 is not negative at $\theta = 0$ because both families CS2 and CS3 are uninformative. Multipoint linkage analysis using markers D13S175, D13S141, and D13S143 gave a maximum lod score of 5.09 at D13S175 with the one-lod-unit support interval spanning 13 centimorgans. These data confirm linkage of the Clouston syndrome locus to 13q11-q12.1. Haplotypes were constructed for all families (**Fig 1**). Linkage data from family CS2 exclude marker D13S143, the closest telomeric flanking marker found by Kibar *et al* (1996). Families CS2 and CS3, both of French–Canadian descent, share alleles (261–103–125–129) at the first four loci (D13S1316-D13S175-D13S141-D13S143, respectively) with the eight French-Canadian families described by Kibar *et al* (1996), supporting a common founder. The CS1 and CS4 families do not share this haplotype, though family CS1 does share the common 125 bp allele for marker D13S141.

We confirm linkage of four ethnically diverse families with Clouston

Manuscript received October 18, 1997; revised February 19, 1998; accepted for publication February 26, 1998.

Reprint requests to: Dr. Susan J. Hayflick, Molecular and Medical Genetics, mail code L103, 3181 SW Sam Jackson Park Road, Portland, OR 97201–3098.

Table I. Combined pairwise lod scores for Clouston syndrome and chromosome 13q markers

	Lod scores (Z) at recombination fraction (q)								
Marker	0.00	0.01	0.05	0.1	0.2	0.3	0.4	q_{max}	Z_{max}
D13S1316	3.61	3.55	3.29	2.95	2.24	1.48	0.71	0.00	3.61
D13S175	2.53	2.48	2.27	1.99	1.39	0.79	0.26	0.00	2.53
D13S141	2.80	2.75	2.56	2.30	1.75	1.17	0.58	0.00	2.80
D13S143	$-\infty$	1.46	1.95	1.96	1.65	1.17	0.60	0.08	1.98
D13S115	$-\infty$	0.65	1.16	1.21	0.99	0.63	0.26	0.08	1.21
D13S1236	4.82	4.73	4.39	3.95	3.01	1.99	0.95	0.00	4.82
D13S1243	$-\infty$	1.21	2.37	2.60	2.28	1.59	0.77	0.10	2.60
D13S221	$-\infty$	−2.06	0.39	1.16	1.42	1.12	0.60	0.18	1.43

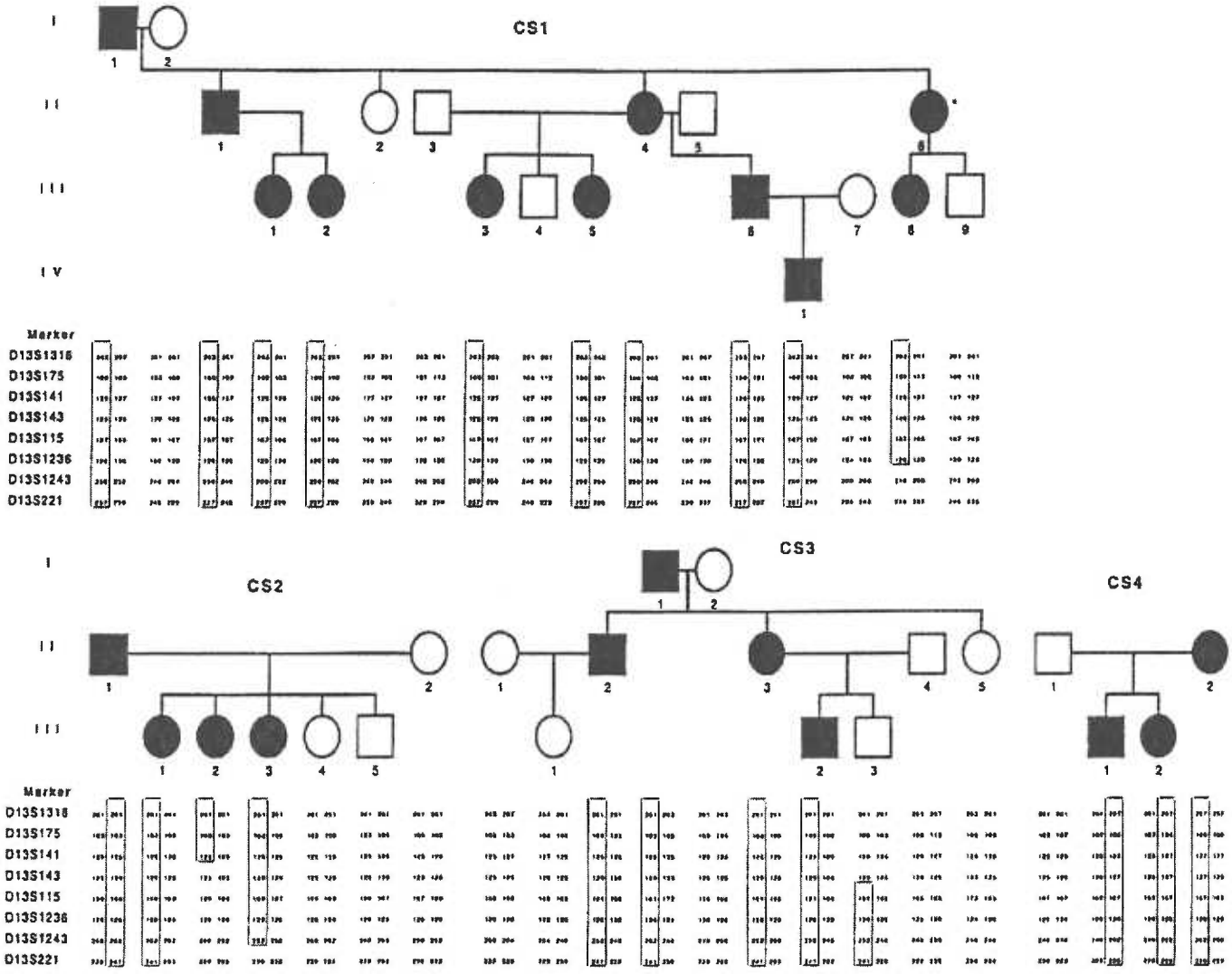


Figure 1. CS family pedigrees and haplotypes. Alleles are given in number of base pairs of the PCR amplification product. Boxes indicate the disease-bearing chromosomal region; solid symbols indicate affected individuals. *DNA not available.

syndrome to chromosome 13q11-q12.1. The two French-Canadian families in this study share a common ancestry with those in the Montreal study and also show a common haplotype. Two families, one of French descent and one of Scottish–Irish descent, do not share this haplotype, suggesting that the mutations in these three groups are of different origin. No clinical features distinguish the common founder (French-Canadian) families from the others.

Our study provides further evidence that Clouston syndrome is a genetically homogeneous disease, which will be important when evaluating candidate genes. Currently, there are no obvious candidate genes in the region. Because the candidate gene interval is large (≈13 centimorgans) additional linkage data narrowing the region will likely be needed before beginning positional cloning.

We thank Dr. Virginia Sybert for informing us about the CS1 family and Drs. Guy Rouleau and Zoha Kibar for sending us reagents. We are grateful to the families and the National Foundation for Ectodermal Dysplasia, whose enthusiastic participation made this project possible. This work was supported by Oregon Health Sciences University Foundation Grant MRF9319 (S.J.H.) and National Institutes of Health Grant R01-DE11311 (J.Z.).

REFERENCES

Clouston HR: A hereditary ectodermal dystrophy. *Can Med Assoc J* 21:18–31, 1929

Clouston HR: The major forms of hereditary ectodermal dysplasia. *Can Med Assoc J* 40:1–7, 1939

Cottingham RW Jr, Idury RM, Schäffer AA: Faster sequential genetic linkage computations. *Am J Hum Genet* 53:252–263, 1993

Dib C, Faure S, Fizames C, *et al*: A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152–154, 1996

Giraud F, Mattei J-F, Rolland M, Ghiglione C, Pommier de Santi P, Sudan N: La dysplasie ectodermique de type Clouston. A propos d'une nouvelle observation avec etude biochimique de la keratine. *Arch Fr Pediatr* 34:982–993, 1977

Guilford P, Dodé C, Crozet F, *et al*: A YAC contig and an EST map in the pericentromeric region of chromosome 13 surrounding the loci for neurosensory nonsyndromic deafness (DFNB1 and DFNA3) and limb-girdle muscular dystrophy type 2C (LGMD2C). *Genomics* 29:163–169, 1995

Gyapay G, Morissette J, Vignal A, *et al*: The 1993–94 Généthon human genetic linkage map. *Nat Genet* 7:246–339, 1994

Hayflick SJ, Taylor T, McKinnon W, Guttmacher AE, Litt M, Zonana J: Clouston syndrome (hidrotic ectodermal dysplasia) is not linked to keratin gene clusters on chromosomes 12 and 17. *J Invest Dermatol* 107:11–14, 1996

Hudson TJ, Engelstein M, Lee MK, *et al*: Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* 13:622–629, 1992

Kibar Z, Der Kaloustian VM, Brais B, Hani V, Fraser FC, Rouleau GA: The gene responsible for Clouston hidrotic ectodermal dysplasia maps to the pericentromeric region of chromosome 13q. *Hum Mol Genet* 5:543–547, 1996

Lathrop GM, Lalouel JM, Julier C, Ott J: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443–3446, 1984

Litt M, Hauge XY, Sharma V: Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. *BioTechniques* 15:280–284, 1993

Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucl Acids Res* 16:1215, 1988

Petrukhin KE, Speer MC, Cayanis E, *et al*: A microsatellite genetic linkage map of human chromosome 13. *Genomics* 15:76–85, 1993

Radhakrishna U, Blouin JL, Mehenni H, *et al*: The gene for autosomal dominant hydrotic ectodermal dysplasia (Clouston syndrome) in a large Indian family maps to the 13q11-q12.1 pericentromeric region. *Am J Hum Genet* 71:80–86, 1997

Schäffer AA: Faster linkage analysis computations for pedigrees with loops or unused alleles. *Hum Hered* 46:226–235, 1996

Schäffer AA, Gupta SK, Shriram K, Cottingham RW Jr: Avoiding recomputation in linkage analysis. *Hum Genet* 44:225–237, 1994

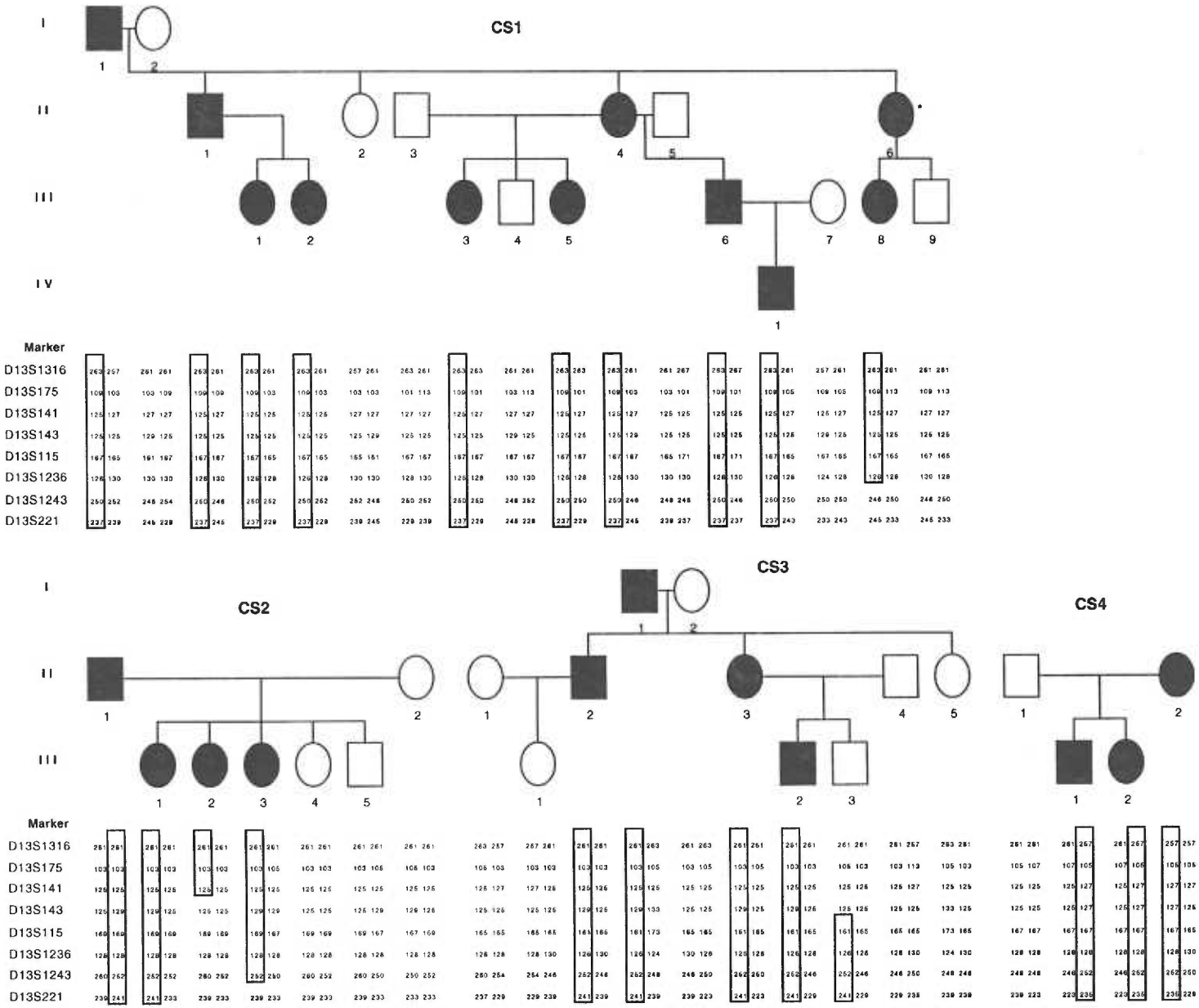


Figure 1. CS family pedigrees and haplotypes. Alleles are given in number of base pairs of the PCR amplification product. Boxes indicate the disease-bearing chromosomal region; solid symbols indicate affected individuals. *DNA not available.