ン」と云う返事が返ってくる様になった，名前の「み」の字を書いてみせ，お名前を書こうかねと促すと名前と読める字を書くようになった，テレビの映像より音楽を好みCDをイヤホンで聞くのを特に楽しむ様になった，以前みられていたように夜中に目を覚まし長時間体を揺することがなくなった，食事後の歯磨き，茶碗おろしを確実にするようになった，関わりの中で笑顔が多くなった，などの変化が起こった．服用後1年3か月後には21年ぶりに字を書くようになり，その後音楽に合わせて童謡などを歌えるようになった．本薬剤投与前の両親の一番の希望は，彼女が歌っている歌をCDに録音したいというものであった．両親の印象によると，これらは初めて出来たことではなく，かつて出来ていたものが少しずつ思い出されるかのように取り返されて来たとのことであった．

患者家族は上述のように塩酸ドネペジル治療の効果について良い印象を保持していたにもかかわらず，それは客観的評価法には必ずしも的確に反映されなかった．行動適応尺度(ABS)，田中ビネー知能検査，絵画語い発達検査，SM社会生活能力検査や田中・ビネー知能検査を使用してみたが，ABSで部分的に効果を認める症例が存在するのみで[43)45)46)]，全例に対応できる評価法はなかった．その後，国際生活機能分類(ICF)をもとに東京学芸大学で作成された心身機能チェックリスト（2006，2007）を使用したところ，知的能力が低いDS患者群で適応しやすいことが判明した．東京学芸大学の菅野らは，このチェックリストを用いてDS患者の日常生活能力の推移を追い，その有用性を確認している[47)]．

重症心身障害児・者施設での二重盲検試験

これまでに確立した用法・用量および評価法を用いて，重症心身障害児・者施設に入所している重症知的障害DS患者に対する24週間の二重盲検試験を行った．同一環境内で過ごすDS患者の日常生活能力を同一の評価者（施設職員）が判定することによって，評価のバラツキやバイアスを可能な限り少なくすることを狙ったものである．

投与開始24週後，実薬群ではプラセボ群と比べて有意差をもって心身機能チェックリストを用いての日常生活能力が高くなっており，特に全般的精神機能が最も上昇し，個別精神機能，尿路機能，音声と発話の機能がそれに続いた．一方，消化器症状については効果を認めなかった[48)]（投稿中）．更に特筆すべきことは，本薬剤投与により，排尿機能障害が著しく改善したことである（投稿準備中）．

本研究の問題点

7年余り本研究を続けているが，その目的は（1）本薬剤がDS患者のQOL向上に有効であるのかどうかを明確にし，(2)有効かつ安全と判断されれば，本薬剤を必要としているDS患者家族への投与に保険適応が降りるように寄与することである．

そのためには，本薬剤の効果を最も的確に反映する客観的な評価法の確立が必要である．我々は，知的障害が重度のDS患者の日常生活能力評価法としては，ICFをもととした心身機能チェックリストが有効ではないかと考えている．しかし，知的発達能力が高いDS患者にこの評価法を用いると，最初から高得点になり治療による効果を認めにくく，適しないと思われる．さらに，知的水準が高いDS患者ではプラセボ効果が高くなることや，DS患者全般で気分的ムラが大きいことなど，DSの特性に即した日常生活能力評価法の確立が必要である．

しかし医療的ニーズが高く，家族の精神面を考慮しても緊急な対応が求められるものは，退行などによって現時点で日常生活能力が非常に低下している患者であろう．そうであれば，現時点でほぼ確立した評価法でも目的を達することが出来るかも知れない．

DSのQOLを高める目的での薬剤治療に関するアンケート調査

「DSは疾病ではなく多様性の1つである」という立場を取り，薬物治療に懐疑的または否定的な人達がいる．その一方，急激退行をきたした子どもを抱えて自身までが精神的諸問題を起こすほどの悲惨な状況がもたらされることもある．そのため，日常生活能力を高める目的での薬剤使用についてDS患者家族がどのように捉えているのかを知るために，アンケート調査を行った．200名中116名（56.0%）から回答があり，結果は概ね支持的であった（表2）．

ま と め

塩酸ドネペジルはADの進行を遅らせることを目的とした治療薬であるが，DSの脳内コリン作動性システムを考慮すると，ADと同様に有効であると推測された．実際に治療に用いた結果からは，効き幅に個人差を認めるものの，おそらくAD患者における改善度を凌駕する有効性があると思われた．このように高い有効性がみられた理由の一つは，DS患者特有の薬物動態にあると思われ，本薬剤の血中濃度がDS患者

表2　ダウン症候群患者家族へのアンケート調査結果

回収率 56.0%

1) 子供の年齢は今幾つか？		a) 0～5歳		b) 6～10歳		c)11～14歳		d) 15歳以上		計	
		23	20.5%	34	30.4%	21	18.8%	34	30.4%	112	100.0%
2) 15歳を越えている方の現状は？(注：重複回答がみられた)	a. ( ) 頃をピークに停滞							1	2.9%	1	10代
	b. 今の所順調							22	64.7%	22	
	c. 言語については問題ない							1	2.9%	1	
	d. 発語聞き取りにくい							13	38.2%	13	
3) 本アンケートの主旨，経緯を読んでいただきましたか？	はい	23	100.0%	33	97.1%	21	100.0%	34	100.0%	111	99.1%
	いいえ		0.0%		0.0%					0	0.0%
	無回答		0.0%	1	2.9%					1	0.9%
4) お子さまが使用する，しないに関わらずお薬自体歓迎するか？	はい	23	100.0%	32	94.1%	19	90.5%	30	88.2%	104	92.9%
	いいえ		0.0%		0.0%	1	4.8%	2	5.9%	3	2.7%
	判断出来ない		0.0%	2	5.9%	1	4.8%	2	5.9%	5	4.5%
5) 保健適応がおりたら，薬の使用を考えるか？	はい	14	60.9%	26	76.5%	16	76.2%	22	64.7%	78	69.6%
	いいえ	3	13.0%	1	2.9%	1	4.8%	5	14.7%	10	8.9%
	わからない	6	26.1%	7	20.6%	4	19.0%	7	20.6%	24	21.4%
6) どれくらいの費用ならば拠出を考えるか？	a. 月千円程度まで	4	17.4%	2	5.9%	0	0.0%		0.0%	6	5.4%
	b. 月3千円程度まで	8	34.8%	16	47.1%	9	42.9%	14	41.2%	47	42.0%
	c. 月5千円程度まで	6	26.1%	8	23.5%	5	23.8%	7	20.6%	26	23.2%
	d. 月1万円程度まで	3	13.0%	3	8.8%	5	23.8%	6	17.6%	17	15.2%
	e. それ以上でも	1	4.3%	3	8.8%	1	4.8%	1	2.9%	6	5.4%
	無回答	1	4.3%	1	2.9%	1	4.8%	6	17.6%	9	8.0%
7) (複数解答可) お子さまの一般の薬についてどのような剤形がのみやすいか	A. 錠剤	2		1		7		15		25	
	B. 小さいサイズの錠剤(または分割出来る錠剤)			10		1		10		21	
	C. ラムネ菓子のように口内で溶ける錠剤	8		11		2		3		24	
	D. 散剤（粉薬）	13		21		10		11		55	
	E. シロップ	8		12		7		6		33	
	F. ゼリーのようなもの	4		10		3		3		20	
	G. 座薬			2						2	
	H. 味は，あまり関係ない	1		4		11		13		29	
	I. 味が良いものがあればよい	13		20		5		5		43	

において健常人より明らかに高いことが効果や副作用の発現に影響を与えていると思われる．そのため，低容量（3mg/日）を基準として，効果が乏しい場合に血中濃度を参考にしつつ増量する方法が安全と思われる．

一方で，DS患者の日常生活能力の変化を捉えるにあたって，現存の評価法の問題点が顕在化した．気分的ムラが大きいDS患者には直接的評価法の再現性は乏しい．種々の客観的評価法は主観的に得られる改善度をうまく反映させえないか，反映できる場合も患者の重症度によっては不適当と考えられた．我々は，ICFに基づく心身機能チェックリストが重症度の高いDS患者の日常生活能力の変化を捉えるのに適していることを見出したが，さらに有用な評価法の確立が急務である．

本薬剤は，認知症の有無やその程度，年齢，IQレベルに関係なくDS治療薬として使用できそうである．DS患者とその家族にとって，日常生活におけるQOLの向上を実現できると大きな福音となる．その反面，自己表現や自己主張ができるようになることが，家族に別のストレスを与えることもあり得る．従って，短絡的に薬物療法のみで問題が解決すると考えるのではなく，環境整備を含めた総合的な取組みを蔑ろにしてはならない．いずれにしても，本薬剤による治療に対するDS家族からの要望は高いので，本薬剤の適応基準や用法用量を厳格に定めた治療ガイドラインを構築していく必要がある．そして何よりも，本薬剤がDS

治療薬としての保険適応を得て，保険診療の元で安心して治療することができるようになることが強く求められる．

結論として，塩酸ドネペジルは，染色体異常症であるDSに対する薬物療法という新しい領域を開く可能性のある薬剤であり，今後の研究がその地位を確固たるものにすることが期待される．

謝辞　本研究は平成14年から開始し，これまでに以下の多くの方々の協力を得て推進することができた．(長崎大学医学部小児科学教室)天本なぎさ先生，土井知己先生，濱田仁美先生，田中温子先生，本田涼子先生；(同泌尿器科)野口　満先生；(同精神神経科)中根秀之先生；(同創薬科学)池田正行教授；(同遺伝情報)柴田義貞先生；(同附属病院薬剤部)中嶋幹郎教授，佐々木　均教授；(長崎大学教育学部)小島道生先生，相川勝代教授，原田純治教授；(長崎大学教育学部附属養護学校)青木瑞恵先生，田中龍彦先生；(東京学芸大学)菅野　敦教授，伊藤　浩先生；(独協医科大学越谷病院小児科)永井敏郎教授；(横浜十愛病院精神神経科)野崎秀次先生；(みさかえの園むつみの家)福田雅文先生；(国立成育医療センター)奥山虎之先生，小崎里華先生，掛江直子先生に深謝いたします．

最後になりましたが，本題をご推薦いただきました日本小児臨床薬理学会の諸先生に御礼申し上げます．

文　献

1) Roizen NJ, Patterson D. Down's syndrome. The Lancet 2003；361：1281—1289.
2) 近藤達郎．Down症候群．小児内科　2009；41(増刊号)：212—215.
3) Kajii T. Predicted prevalence of Down syndormelive births in Japan, 1970-2006. Am J Med Genet 2008；146A：1387—1388.
4) Glasson EJ, Sullivan SG, Hussain R, et al. The changing survival profile of people with Down's syndrome：Implications for genetic counseling. Clin Genet 2002；62：390—393.
5) Rumble B, Retallack R, Hilbich C, et al. Amyloid A4 protein and its precursor in Down's syndrome and Alzheimer's disease. N Engl J Med 1989；320：1446—1452.
6) Yates CM, Simpson J, Maloney AF, et al. Alzheimer-like cholinergic deficiency in Down syndrome. Lancet 1980；2：979.
7) Casanova MF, Walker LC, Whitehouse PJ, et al. Abnormalities of the nucleus basalis in Down syndrome. Ann Neurol 1985；18：310—313.
8) Mann DM, Esiri MM. The pattern of acquisition of plaques and tangles in the brains of patients under 50 years of age with Down's syndrome. J Neurol Sci 1989；89：169—179.
9) 菅野　敦．退行を示した青年期・成人期知的障害者に対する地域生活支援と社会参加の促進に関する研究—退行の類型と予防—．発達障害支援システム学研究　2005；4：35—46.
10) 菅野　敦．成人期に生じた問題への支援．菅野敦，池田由紀江編．ダウン症者の豊かな生活．東京：福村出版，1998：82—124.
11) Capone G, Goyal P, Ares W, et al. Neyrobehavioral disorders in childres, adolescents, and young adults with Down syndrome. Am J Med Genet. 2006；142C：158—172.
12) Ball SL, Holland AJ, Hon J, et al. Personality and behavior changes mark the early stages of Alzheimer's disease in adults with Down's syndrome：findings from a prospective population-based study. International Journal of Geriatric Psychiatry 2006；21：661—673.
13) Cummings JL, McRae T, Zhang R；Donepezil-Sertraline Study Group. Effects of donepezil on neuropsychiatric symptoms in patients with dementia and severe behavioral disoprders. Am J Geriatr Psychiatry 2006；14：605—612.
14) Wisniewski KE, Willims HJ, Wisniewski HM. Precocious aging and dementia in patients with Down's syndrome. Biol Psychiatry 1978；123：619—627.
15) Lai F, Williams RS. A prospective study of Altzheimer disease in Down's syndrome. Arch Neurol 1989；6：849—853.
16) 高嶋幸男．ダウン症候群とアルツハイマー型痴呆．発達障害研究　1990；12：50—53.
17) Wisniewski KE, Wisniewski HM, When GY. Occurrence of neuropathological changes and dementia of Alzheimer's disease in Down's syndrome. Ann Neurol 1985；17：278—282.
18) Kish SJ, Distefano LM, Dozic S, et al. [3H] vesamicol binding in human brain cholinergic deficiency disprders. Neurosci Lett 1990；117：347—352.
19) Yates CM, Simpson J, Gordon A, et al. Catecholamines and cholinergic enzymes in pre-senile and senile Alzheimer-type dementia and Down's syndrome. Brain Res 1983；280：119—126.
20) Casanova MF, Walker LC, Whitehouse PJ, et al. Abnormalities of the nucleus basalis in Down's syndrome. Ann Neurol 1985；18：310—313.
21) McGeer EG, Norman M, Boyes B, et al. Acetylcholine and aromatic amine systems in postmortem brain of the infant with Down's syndrome. Exp Neurol 1985；87：557—570.
22) Florez J, del Arco C, Gonzalez A, et al. Autoradiographic studies of neurotransmitter receptors in the brain of newborn infants with Down syndrome. Am J Med Genet 1990；7(Suppl)：301—305.
23) Holtzman DM, Santucci D, Kilbridge J, et al. Developmental abnormalities and age-related neurodegeneration in a mouse model of Down syndrome. ProcNatl Aca Sci USA 1996；93：13333—13338.
24) Hunter CL, Bimonte HA, Granholm AC. Behavioral comparison of 4 and 6 month-old Ts65Dn mice：age-related impairments in working and reference memory. Behav Brain Res 2003；138：121—131.

25) Sacks B, Smith S. Peaple with Down's syndrome-can be distinguished on the basis of cholinergic dysfunction. J Neurol Neurosurg Psychiatry 1989；52：1294—1295.
26) Beccaria L, Merziani E, Manzoni P, et al. Further evidence of cholinergic impairment of the neuroendocrine control of the GH secretion in Down's syndrome. Dement Geriatr Cogn Disord 1998；9：78—81.
27) Hardy JA, Higgins GA. Alzheimer's disease：The amyloid cascade hypothesis. Science 1992；256：184—185.
28) Salehi A, Delcroix JD, Belichenko PV, et al. Increased App expression in a mouse model of Down's syndrome disrupts NGF transport and cause cholinergic neuron degeneration. Neuron 2006；51：29—42.
29) Raux G, Guyant-Marechal L, Martin C, et al. Molecular diagnosis of autosomal dominant early onset Alzheimer's disease：an update. J Med Genet 2005；42：793—795.
30) Theuns J, Brouwers N, Engelborghs S, et al. Promoter mutations that increase amyloid precursor-protein expression are associated with Alzheimer disease. Am J Hum Genet 2006；78：936—946.
31) Kishnani PS, Sullivan JA, Walter BK, et al. Cholinergic therapy for Down's syndrome. Lancet 1999；353：1064—1065.
32) Hemingway-Eltomey JM, Lerner AJ. Adverse effects of donepezil in treating Alzheimer's disease associated with Down's sydnrome. Am J Psychiatry 1999；156：1470.
33) Lott IT, Osann K, Doran E, et al. Down syndrome and Alzheimer disease：response to donepezil. Arch Neurol 2002；59：1133—1136.
34) Heller JH, Spiridigliozzi GA, Sullivan JA, et al. Donepezil for the treatment of language deficits in aduls with Down syndrome：a preliminary 24-week open trial. Am J Med Genet 2003；116A：111—116.
35) Cipriani G, Bianchetti A, Trabucchi M. Donepezil use in the treatment of dementia associated with Down syndrome. Arch Neurol 2003；60：292.
36) Heller JH, Spiridigliozzi GA, Doraiswamy PM, et al. Donepezil effects on language in children with Down syndrome：results of the first 22-week pilot clinical trial. Am J Med Genet A 2004；130：325—326.
37) Kishnani PS, Sommer BR, Handen BL, et al. The efficacy, safety, and tolerability of donepezil for the treatment of young adult with Down syndome. Am J Med Genet 2009；149A：1641—1654.
38) 土井知己，近藤達郎，天本なぎさ，他．ダウン症候群の長崎県における現状—父母へのアンケートを中心として—．長崎県医師会報　1998；631：3—11.
39) 土井知己，近藤達郎，天本なぎさ，他．ダウン症候群の長崎県における現状—父母へのアンケートを中心として—．小児科臨床　1999；52：67—72.
40) 土井知己，近藤達郎，船越康智．ダウン症候群児・者を取り巻く環境の変移．日本遺伝カウンセリング学会誌　2004；25：81—87.
41) 近藤達郎．地域のダウン症候群患者の自立支援のための医療連携．小児歯科臨床　2008；12：12—17.
42) Kondoh T, Nakashima M, Sasaki H, et al. Pharmacokinetics of donepezil in Down syndrome. Ann Pharmacother 2005；39：572—573.
43) 近藤達郎，森内浩幸．ダウン症候群患者のQOL向上のための塩酸ドネペジル療法．小児内科　2009；41：916—918.
44) 近藤達郎．ダウン症候群患者における日常生活能力改善のための塩酸ドネペジル投与に関する研究．日本小児臨床薬理学会誌　2006；19：123—127.
45) Kondoh T, Amamoto N, Doi T, et al. Dramatic Improvement in Down Syndrome-Associated Cognitive Impairment with Donepezil. Ann Pharmacother 2005；39：563—566.
46) Kondoh T, Amamoto N, Doi T, et al. Dramatic Improvement on donepezil in Down syndrome-Associated Cognitive Impairment. Dementia, Helix review series 2006；8：6—9.
47) 伊藤　浩，菅野　敦．知的障害者の退行・早期老化の評価尺度としての心身機能チェックリストの有効性に関する研究．発達障害支援システム学研究　2008；7：9—17.
48) ドネペジル療法でダウン症候群のQOL改善効果を確認．Medical Tribune 2009年3月12日号．2009：21.

European Journal of Medical Genetics 53 (2010) 244–249

Contents lists available at ScienceDirect

European Journal of Medical Genetics

journal homepage: http://www.elsevier.com/locate/ejmg

Original article

Familial brain arteriovenous malformation maps to 5p13–q14, 15q11–q13 or 18p11: Linkage analysis with clipped fingernail DNA on high-density SNP array

Masahiro Oikawa [a,b], Hideo Kuniba [a,c], Tatsuro Kondoh [d], Akira Kinoshita [a,f], Takeshi Nagayasu [b], Norio Niikawa [e,f], Koh-ichiro Yoshiura [a,f,*]

[a] Departments of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan
[b] Departments of Surgical Oncology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan
[c] Departments of Pediatrics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan
[d] Division for Developmental Disabilities, The Mutsumi House, Misakaenosono Institute for Persons with Severe Intellectural/Motor Disabilities, Konagai-Cho, Japan
[e] Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Tobetsu, Japan
[f] Solution Oriented Research for Science and Technology (SORST), Japan Science and Technology Agency (JST), Tokyo, Japan

ARTICLE INFO

Article history:
Received 12 February 2010
Accepted 11 June 2010
Available online 22 June 2010

Keywords:
Arteriovenous malformation
Genomewide linkage analysis
Fingernail DNA
Mutation search
GeneChip™

ABSTRACT

Familial arteriovenous malformations (AVM) in the brain is a very rare disease. It is defined as its occurrence in two or more relatives (up to third-degree relatives) in a family without any associated disorders, such as hereditary hemorrhagic telangiectasia. We encountered a Japanese family with brain AVM in which four affected members in four successive generations were observed. One DNA sample extracted from leukocytes of the proband and ten DNA samples from clipped finger nails of other members were available. A genome-wide linkage analysis was performed on this pedigree using Affymetrix GeneCip 10K 2.0 Xba Array and MERLIN software. We obtained sufficient performance of SNP genotyping in the fingernail samples with the mean SNP call rate of 92.49%, and identified 18 regions with positive LOD scores. Haplotype and linkage analyses with microsatellite markers at these regions confirmed three possible disease-responsible regions, i.e., 5p13.2–q14.1, 15q11.2–q13.1 and 18p11.32–p11.22. Sequence analysis was conducted for ten selected candidate genes at 5p13.2–q14.1, such as *MAP3K1*, *DAB2*, *OCLN*, *FGF10*, *ESM1*, *ITGA1*, *ITGA2*, *EGFLAM*, *ERBB2IP*, and *PIK3R1*, but no causative genetic alteration was detected. This is the first experience of adoption of fingernail DNA to genome-wide, high-density SNP microarray analysis, showing candidate brain AVM susceptible regions.

1. Introduction

Arteriovenous malformation (AVM) in the brain is a disease defined by the presence of arteriovenous shunt(s) through a nidus of coiled and tortuous vascular connections between feeding arteries and draining veins within the brain parenchyma [10]. This vascular malformation is thought to be congenital, and develops before or after birth [7] from a residual of the primitive artery–vein connection. Its most common symptom is intracranial hemorrhage with an estimated risk of 1.3–3.9% yearly after the diagnosis of AVM [4]. Other signs may include intractable seizures, headache and ischemic steal syndrome. The prevalence of AVM is estimated to be approximately 0.01% and the detection rate ranges between 1.12 and 1.34 per 100,000 person years [7,10]. Although most cases of AVM are sporadic, a total of 53 patients from 25 families have been reported [27]. Familial brain AVM is defined when it occurs in two or more relatives (up to third-degree relative) in a family without associated disorders such as hereditary hemorrhagic telangiectasia (HHT), is autosomal dominant multisystemic vascular dysplasia [9,27]. It is plausible that familial cases are more frequent and could be overlooked because of asymptomatic conditions in other relatives.

Although several causative genes have been elucidated in some heritable syndromic AVM [2,3,5,6,12,13,17,20,21,23,26], molecular genetic studies of familial or sporadic AVM remain scant. HHT type 1 (HHT1) and HHT type 2 (HHT2) are known to be caused by mutations in *ENG* at 9q34.11 and *ACVRL1* (or *ALK1*) at 12q13.13, respectively [12,17]. Mutations in *RASA1* at 5q14.3 cause capillary malformation-arteriovenous malformation (CM-AVM) [3,6,20,21,26] characterized by small, round-to-oval, pink-red and multiple CM: one-third of CV-AVM patients also has fast-flow lesions such as AVM. Mutations in *PTEN* have been implicated in PTEN hamartoma tumor syndromes including Bannayan–Riley–Ruvalcaba syndrome, in which AVM

* Corresponding author. Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Sakamoto 1-12-4, Nagasaki 852-8523, Japan. Tel.: +81 95 819 7118; fax: +81 95 819 7121.
E-mail address: kyoshi@nagasaki-u.ac.jp (K. Yoshiura).

1769-7212/$ – see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmg.2010.06.007

occasionally presents [23]. Three genes, *KIRIT1* (*CCM1*) [13] at 7q21.2, *MGC4607* (*CCM2*) [5] at 7p13 and *PDCD10* (*CCM3*) [2] at 3q26.1, are responsible for cerebral cavernous malformation (hamartomatous vascular malformations). On the other hand, regarding familial AVM, only two linkage analyses using 6 small families have been published by a research group [11,25], showing seven possible disease-responsible regions, i.e., 6q25 with the highest LOD score, 3p27, 4q34, 7p21, 13q32–q33, 16p13–q12 and 20q11–q13, but failed to identify the causative mutation. In sporadic brain AVM, microarray study showed that the *VEGFA*, *ITGA5*, *ENG* and *MMP9* genes that may involve vascular development or maintenance, are highly expressed in AVM compared with normal brain parenchyma [8,22,24].

Here we report results of a genome-wide linkage analysis on an AVM family with four affected members in two successive generations.

2. Materials and methods

2.1. Subjects

A Japanese family consisting of 19 members across four generations included two patients with brain AVM, one patient with pulmonary AVM and one patient with both brain and pulmonary AVM (Fig. 1). The proband (III-3) first exhibited intractable epilepsy at 13 years old and was diagnosed by magnetic resonance imaging (MRI) as having a brain AVM of 2 cm in diameter located in the right frontal lobe (Fig. 2). Chest X-ray at the first visit detected a nodular shadow in the right lower lung field, and a diagnoses of pulmonary AVM with a 24% of shunt-rate was made following angiogram made (Fig. 2). This was resected when the proband was 14 years old. The proband's brain AVM was treated by gamma knife surgery when she was 19 years old, followed by treatment with antiepileptic medication. Her mother (II-3) died of intracranial hemorrhage due to brain AVM, and the maternal grandfather (I-1) died of a cancer. Another patient (III-5) had asymptomatic brain AVM, which was accidentally diagnosed by MRI. His father (II-5) had pulmonary AVM instead of brain AVM. These four members were assigned to "affected", six members (II-6, III-1, III-6, III-7, IV-1, and IV-2) without AVM confirmed by MRI were "unaffected", and the remaining three (I-2, II-1, and IV-3) who were not assessed by MRI but had neither past history of recurrent epistaxis or gastrointestinal tract bleedings were "unknown". None of the members had any AVM-related diseases, such as HHT. Evaluation of cutaneous

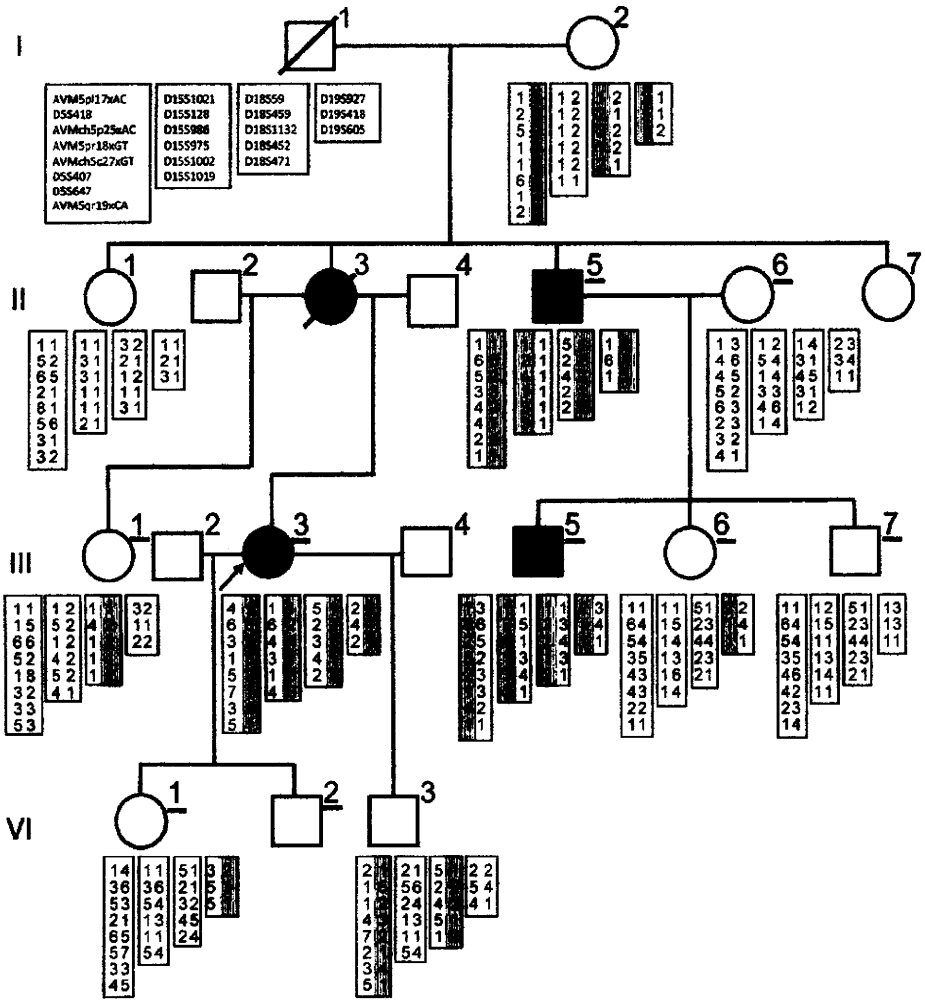

Fig. 1. Results of haplotype analysis at polymorphic loci in four regions, 5q13.2–q14.1, 15q11.2–q13.1, 18p11.32–p11.22 and 19q13.3–q13.42. Underlined individuals indicate those examined by MRI, and DNA was unavailable from individuals without haplotypes. Polymorphic alleles are numbered and candidate disease-associated haplotypes are shown by dotted boxes. Primer sequences designed for CA repeat amplification are available in Supplementary Table.

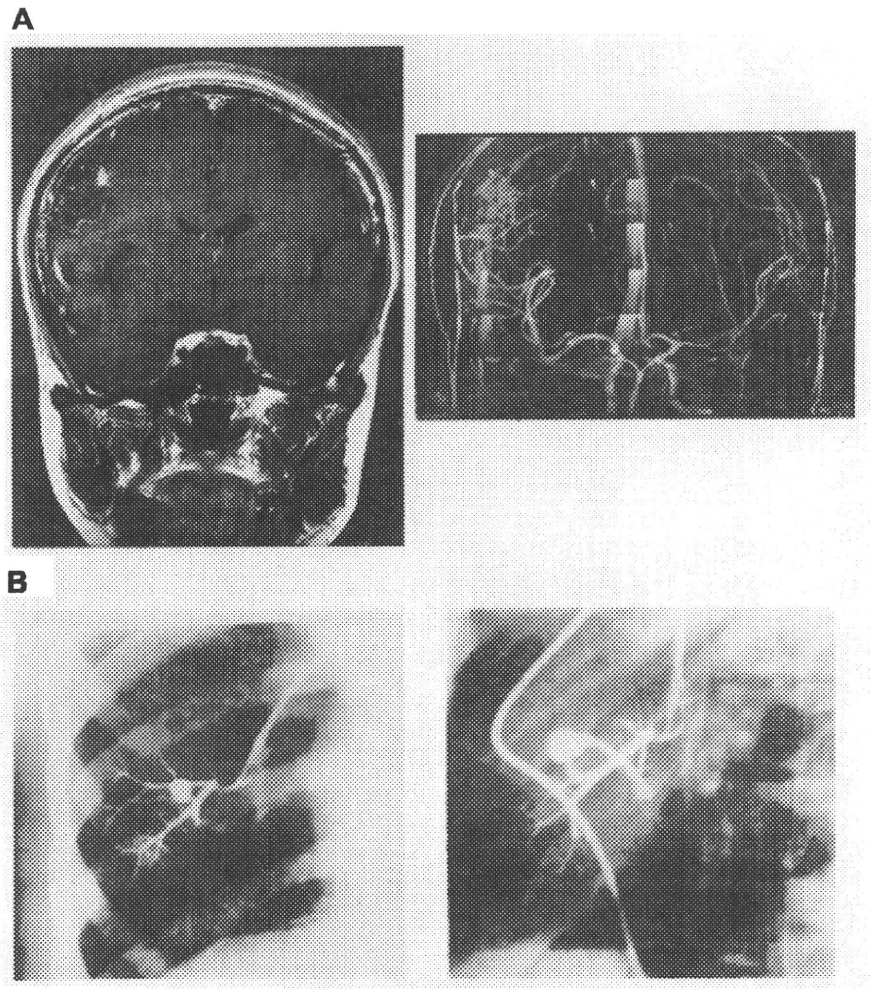

Fig. 2. Imaging of the brain and pulmonary AVM in the proband. (A) MRI scan and MR angiogram of the proband. The AVM is located right frontal lobe measured 2.0 × 1.3 cm. (B) Pulmonary angiograms of the proband. The pulmonary AVM is located in the right lower lobe (rtS8b) with 24% of shunt-rate.

lesions was conducted by examination of the proband and by detailed interview of the other family members by the proband and her sister (III-1), who is nurse. A total of 13 members participated in this study under informed consent. All experimental procedures for this study were approved by Committee for the Ethical Issues on Human Genome and Gene Analysis in Nagasaki University.

2.2. DNA extraction

As a blood sample was available only from the proband, clipped fingernail samples were obtained from 10 of the other 12 members instead. Genomic DNA was extracted from the fingernails using a buffer solution containing urea, DDT and proteinase K, as reported previously [16,18]. Briefly, clipped fingernails were once frozen in liquid nitrogen and crushed into fine powder using Multi-beads Shocker™ (Yasui Kikai, Osaka, Japan). The nail powder was lysed in a urea-lysis solution (2 M urea; 0.5% SDS; 10 mM Tris–HCl, pH 7.5; 0.1 M EDTA) containing 1 mg/ml proteinase K and 40 mM DDT at 55 °C overnight. Nail DNA was extracted with phenol/chloroform, and precipitated with ethanol and sodium acetate. Precipitated nail DNA was dissolved again in extraction buffer (0.5% SDS; 10 mM Tris–HCl, pH 7.5; 0.1 M EDTA) containing 1 mg/ml proteinase K, and incubated at 55 °C overnight. DNA was purified again as above, and was suspended in 30 μl of 1× TE buffer.

2.3. SNP genotyping with Affymetrix 10K 2.0 array

Blood DNA (250 ng) was processed according to the standard protocol provided by the GeneChip Mapping 10K Xba Assay Kit (Affymetrix, Santa Clara, CA). Fingernail DNA was processed in a similar manner but with the two following modifications to adapt to the oligonucleotide microarray system [15]. Prolongation of digestion time from 120 min as the standard protocol to overnight; and increase of the PCR cycle number from 35 to 45 cycles. Data acquired from the Affymetrix GeneChip Operating System were analyzed using the Affymetrix GeneChip Genotyping Analysis Software (GTYPE) 4.0 to call genotypes.

2.4. Linkage analysis with SNP-genotype data and haplotype analysis with microsatellite markers

Multipoint LOD scores were calculated using MERLIN software [1], under an assumption that AVM in the family is transmitted in an autosomal dominant mode with reduced penetrance ($p = 0.9$)

and with the disease allele frequency of 0.001. At loci with a positive LOD score by the GeneChip genotyping, possibly disease-associated haplotypes were constructed using SNP calls.

When SNP information was not informative, microsatellite markers were used for genotyping. Microsatellite markers used were referred to the National Center for Biotechnology Information (NCBI) database. One each of primer pairs for the markers was labeled with FAM, HEX, or NED (Supplementary Table 1), and PCR was performed in a 10 μl mixture containing 5 ng genomic DNA; 0.25 U ExTaq DNA polymerase HS-version (TAKARA Bio Inc., Kyoto, Japan); 200 μM dNTP; 0.5 μM primer; 1× ExTaq buffer on the T1 Thermocycler (Biometra, Goettingen, Germany). PCR products were separated on Genetic Analyzer 3130xl (AppliedBiosystems), and genotyping was carried out using GeneMapper software (AppliedBiosystems). At the regions where the affected individuals have a disease-associated haplotype, two-point LOD score was calculated by MLINK program (included in FASTLINK software version 4.0P) [14].

2.5. Mutation analysis

Some genes located within candidate regions identified by the linkage analysis were selected for further mutation analysis. A few other genes, albeit outside the regions, were also subjected to mutation analysis. Primer pairs for such genes were designed using Primer3-web 0.3.0 (http://frodo.wi.mit.edu/primer3/input.htm), according to their genomic sequences retrieved from the University of California, Santa Cruz (UCSC) Genome Browser Home (http://genome.cse.ucsc.edu/). PCR was carried out in a 15 μl reaction mixture containing 5 ng DNA; 0.25 U ExTaq DNA polymerase HS version; 200 mM dNTP; 0.5 μM each primer; 1× ExTaq buffer on the T1 Thermocycler. PCR products were subjected to direct sequencing, using BigDye Terminator v3.1 Cycle sequencing Kit (AppliedBiosystems) and Genetic Analyzer 3130xl. Electropherograms of sequences were aligned with ATGC software (GENETYX Corp., Tokyo, Japan) to inspect base alterations.

2.6. Search for genomic aberration

To search for copy number change within the candidate loci identified by linkage analysis, we used Affymetrix® Genome-Wide Human SNP Array 5.0 (920,568 probes; Affymetrix). Genomic DNA extracted from white blood cell of proband was processed according to manufacture's protocol. Intensity data from each probes were obtained from Affymetrix® Genotyping Console 3.0 as a CEL files. Unpaired copy number analysis of whole genome was carried out using Partek Genomics Suite (Partek, MO, USA) and regions with copy number change were determined by Hidden Markov Model at default settings.

3. Results

3.1. Linkage and haplotype analyses

The mean SNP call rate was 92.49% in 11 fingernail DNA samples, compared to 98.11% in a blood DNA sample from the proband. Incorrect SNP calls may result in seemingly inconsistent parent–child transmissions, but the call rates obtained are actually enough for further studies. We thus advanced to calculate LOD scores using these data.

The linkage analysis using MERLIN software revealed 18 regions with positive LOD scores (>0.00). Of the 18 regions, 14 with the following conditions were excluded: those without any functional full-length RefSeq genes; those in small size (<200 kb); and those in which some affected members did not have a common haplotype. Consequently, four loci, 5p13.2–q14.1, 15q11.2–q13.1, 18p11.32–p11.22 and 19q13.33–q13.42, remained as possibly linked regions (Figs. 1 and 3).

We then genotyped with microsatellite markers and calculated two-point LOD scores, considering the affected, unaffected, and the unknown family members. We confirmed three of the four candidate loci. They were a 48-Mb region between markers rs1366265 and rs1373965 at 5p13.2–q14.1, a 6-Mb region between rs850819

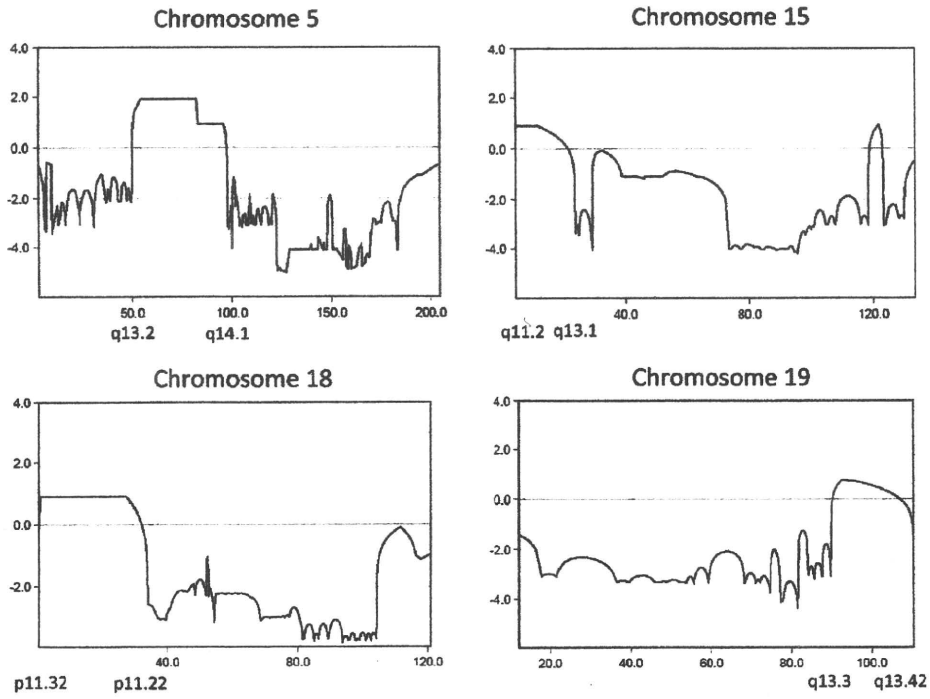

Fig. 3. Multipoint LOD scores calculated by MERLIN in four chromosomal regions, 5q13.2–q14.1, 15q11.2–q13.1, 18p11.32–p11.22 and 19q13.3–q13.42.

and rs818089 at 15q11.2–q13.1, both giving the maximum two-point LOD score of 1.632 ($\theta = 0$), and a 9-Mb region between rs486633 and rs1942150 at 18p11.32–p11.22 with the maximum LOD score of 0.851 ($\theta = 0$) (Table 1). As a possibly disease-associated haplotype on 19q13.33–q13.42 was transmitted to two definitively unaffected individuals (III-6 and IV-1), chromosome 19 was ruled out from the candidacy (Table 1, Fig. 1).

3.2. Mutation analysis of candidate genes

Within the 48-Mb region at 5p13.2–q14.1, there are about 200 RefSeq genes. Ten (*MAP3K1, DAB2, OCLN, FGF10, ESM1, ITGA1, ITGA2, EDFLAM, ERBB2IP,* and *PIK3R1*) from these genes were focused and selected as candidates for brain AVM, since they concern development or maintenance of vessels, are associated with other heritable vascular disorders such as HHT, or are expressed in the brain with AVM [8,22,24]. Mutation analyses in these 10 genes revealed no pathologic mutation in the proband, although other affected members were not examined because of insufficient amount of their DNA. Although the genes endoglin isoform 1 precursor (*ENG*), activin A receptor type II like 1 (*ALK1*) and RAS p21 protein activator 1 (*RASA1*) are not located in the candidate region, we investigated whether any of them are involved in the etiology of AVM in the family as a partial symptom of HHT or AVM-CM. Direct sequencing of these three genes failed to show any causative variants.

Copy number analysis of proband revealed one increased copy number loci at 12q and decreased at 2p, 3q, 4q, 6p, 7q and 22q (data not shown). But all these alterations were reported previously as copy number polymorphisms (http://projects.tcag.ca/variation/) and out of our candidate loci. In addition, neither deletions nor microdeletions were detected at 9q34.11 of *ENG*, 12q13.13 of *ALK1* and 5q14.3 of *RASA1*.

4. Discussion

We have reported a family consisting of two affected members with brain AVM, one with pulmonary AVM and one with both brain and pulmonary AVM. The condition in this family met the criteria of familial brain AVM and seems to be inherited in an autosomal dominant mode. We tried to assign the location of a putative disease-gene by linkage analysis and search for mutations by subsequent candidate gene approach.

The linkage analysis of the family revealed three candidate regions (5p13.2–q14.1, 15q11.2–q13.1, and 18p11.32–p11.22) with relatively high LOD scores of 1.632, 1.632 and 0.851, respectively (Table 1). However, neither region was conclusive. This insufficient mapping may have arisen from the small pedigree size, and/or from incomplete ascertainment of affected members, e.g., probable existence of asymptomatic affected persons among the "unknown" members. Indeed, as for a candidate locus at 5p13.2–q14.1, the proband's maternal grandmother (I-2) and son (IV-3) had a haplotype common to the three affected members (Fig. 1), but they were fallen into the "unknown" individuals. If DNA from IV-2 was available and if MRI examinations of VI-3 and I-2 were carried out, we would have obtained more definitive results. As we performed linkage analysis using high-density SNP genotyping, 14 small regions not containing RefSeq genes or miRNAs showed a positive LOD score. It is possible that an unidentified transcribed RNA in one of these regions could cause familial AVM, but these regions are candidate loci with a lower priority than those containing known genes. Thus, the three regions have remained at present as the equally possible loci for AVM. The three regions do not overlap with a previously reported candidate locus of familial brain AVM, i.e., 6p25 [11], and do not contain genes responsible for syndromic AVM (heritable disorders involving AVM) or cerebral cavernous malformations, such as *ENG* [17], *ALK1* [12], *RASA1* [3,6,20,21,26], and *PTEN* [23], *KRIT1* [13], *MGC407* [5], *PDCD10* [2].

We then searched for mutations in 10 genes within 5p13.2–q14.1, among which *MAP3K1*, *DAB2* and *OCLN* encode proteins playing roles in the TGF-β signaling pathway, and *FGF10, ESM1, ITGA1, ITGA2, EGFLAM, ERBB2IP* and *PIK3R1* were those expressed in brain AVM tissues by previous microarray analysis [8,22,24]. Nevertheless, no pathologic mutation was found in any of them. Because the presence of both brain AVM and pulmonary AVM in this pedigree is reminiscent of Hereditary Hemorrhagic Telangiectasia, we analyzed *ENG* and *ALK1* for mutations and genomic aberrations, which may cause HHT1 and HHT2 respectively [12,17]. The proband did not have any mutations in the coding exons or intron/exon boundaries of either gene, nor any genomic aberrations at those loci. We also analyzed *RASA1* because this may cause CM-AVM, which is characterized by multiple CM and AVM [3,6,20,21,26]. No causative mutation or genomic aberration was detected in the proband. Although other genes, such as *KRIT1, MGC407* and *PDCD10*, have been shown to cause slow-flow lesions i.e., cerebral cavernous malformation [2,5,13], they were not investigated in the present study, because the clinical manifestations in our family did not meet the criteria for these diseases.

Participation of family members and compliance with guidelines for human genome researches are critical to conduct a linkage analysis. Whole-blood samples cannot occasionally be available in some family members because of their far domicile. In such the case, fingernail DNA is useful, since clipped fingernails can be mailed in a usual way, and stored long at a room temperature, as indicated previously [16,19]. The present study is the first experience to adopt fingernail DNA to genome-wide high-density SNP microarray analysis. The performance obtained from fingernail DNA was sufficient, showing all SNP call rates of >86%. According to the manufacture's protocol, samples with an SNP call rate of <85% should further be evaluated before including the data in downstream analysis. Incorrect SNP calls may make serious problems in linkage analysis. For instance, SNPs with parent–child transmission inconsistency may be omitted, leading to a reduced LOD score.

In conclusion, we have assigned the familial AVM locus to three alternative regions, 5p13.2–q13.2, 15q11.2–q13.1 and 18p11.32–p11.22, by a genome-wide, high-density, SNP-based

Table 1
Two-point LOD scores for brain AVM at various loci.

Locus	Recombination fraction (θ)					
	0.00	0.01	0.02	0.03	0.04	0.05
AVM5pl17xAC	0.032	0.030	0.029	0.027	0.026	0.025
D5S418	0.551	0.535	0.518	0.501	0.484	0.467
AVMch5p25xAC	1.334	1.301	1.268	1.234	1.201	1.167
AVM5pr18xGT	0.511	0.491	0.472	0.452	0.433	0.414
AVMch5c27xGT	1.630	1.597	1.564	1.531	1.497	1.463
AVM5c18xAC	1.373	1.344	1.314	1.285	1.255	1.225
D5S407	1.632	1.599	1.566	1.532	1.499	1.465
D5S647	1.154	1.121	1.089	1.056	1.023	0.991
AVM5qr19xCA	0.810	0.790	0.769	0.748	0.727	0.706
D15S1021	0.171	0.164	0.157	0.150	0.143	0.137
D15S128	0.876	0.858	0.841	0.823	0.805	0.787
D15S986	0.812	0.791	0.770	0.749	0.728	0.707
D15S975	0.400	0.387	0.374	0.361	0.348	0.335
D15S1002	1.330	1.298	1.266	1.234	1.202	1.170
D15S1019	1.632	1.599	1.566	1.532	1.499	1.465
D18S59	0.199	0.214	0.225	0.234	0.241	0.246
D18S459	0.142	0.136	0.131	0.125	0.120	0.114
D18S1132	0.677	0.663	0.650	0.636	0.623	0.609
D18S452	0.851	0.832	0.813	0.794	0.774	0.755
D18S471	0.240	0.231	0.222	0.214	0.205	0.197
D19S927	−0.302	−0.277	−0.254	−0.234	−0.216	−0.200
D19S418	−2.655	−2.453	−2.257	−2.078	−1.919	−1.778
D19S605	−0.648	−0.574	−0.512	−0.460	−0.414	−0.374

linkage analysis with fingernail DNA in an AVM family. However, mutation analyses of some genes in the regions failed to identify any pathological changes.

Funding

This research was supported in part by Grants-in-Aid for Scientific Research (Nos. 17019055 and 19390095) from the Ministry of Education, Sports, Culture, Science and Technology of Japan, and by SORST from Japan Science and Technology Agency (JST). K.Y. was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Health, Labour and Welfare of Japan.

Competing interests

There are no competing interests.

Acknowledgments

We are grateful to the family members for their participation in this research. We also thank Ms Yasuko Noguchi, Ms Miho Ooga, and Ms Chisa Hayashida for their technical assistance.

Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmg.2010.06.007.

References

[1] G.R. Abecasis, S.S. Cherny, W.O. Cookson, L.R. Cardon, Merlin—rapid analysis of dense genetic maps using sparse gene flow trees, Nat. Genet. 30 (2002) 97–101.
[2] F. Bergametti, C. Denier, P. Labauge, M. Arnoult, S. Boetto, M. Clanet, P. Coubes, B. Echenne, R. Ibrahim, B. Irthum, G. Jacquet, M. Lonjon, J.J. Moreau, J.P. Neau, F. Parker, M. Tremoulet, E. Tournier-Lasserve, Société Française de Neurochirurgie, Mutations within the programmed cell death 10 gene cause cerebral cavernous malformations, Am. J. Hum. Genet. 76 (2005) 42–51.
[3] L.M. Boon, J.B. Mulliken, M. Vikkula, RASA1: variable phenotype with capillary and arteriovenous malformation, Curr. Opin. Genet. Dev. 15 (2005) 265–269.
[4] P.M. Crawford, C.R. West, D.W. Chadwick, M.D.M. Shaw, Arteriovenous malformations of the brain: natural history in unoperated patients, J. Neurol. Neurosurg. Psychiatr. 49 (1986) 1–10.
[5] C. Denier, S. Goutagny, P. Labauge, V. Krivosic, M. Arnoult, A. Cousin, A.L. Benabid, J. Comoy, P. Frerebeau, B. Gilbert, J.P. Houtteville, M. Jan, F. Lapierre, H. Loiseau, P. Menei, P. Mercier, J.J. Moreau, A. Nivelon-Chevallier, F. Parker, A.M. Redondo, J.M. Scarabin, M. Tremoulet, M. Zerah, J. Maciazek, E. Tournier-Lasserve, Société Française de Neurochirurgie, Mutations within the MGC4607 gene cause cerebral cavernous malformations, Am. J. Hum. Genet. 74 (2004) 326–337.
[6] I. Eerola, L.M. Boon, J.B. Mulliken, P.E. Burrows, A. Dompmartin, S. Watanabe, R. Vanwijck, M. Vikkula, Capillary malformation-arteriovenous malformation, a new clinical and genetic disorder caused by RASA1 mutations, Am. J. Hum. Genet. 73 (2003) 1240–1249.
[7] I.G. Fleetwood, G.K. Steinberg, Arteriovenous malformations, Lancet 359 (2002) 863–873.
[8] T. Hashimoto, M.T. Lawton, G. Wen, G.Y. Yang, T. Chaly, C.L. Stewart, H.K. Dressman, N.M. Barbaro, D.A. Marchuk, W.L. Young, Gene microarray analysis of human brain arteriovenous malformations, Neurosurgery 54 (2004) 410–423 discussion 423–425.
[9] R. Herzig, S. Burval, V. Vladyke, L. Janouskova, P. Krivanek, B. Krupka, I. Vlachova, K. Urbanek, Familial occurrence of cerebral arteriovenous malformation in sisters: case report and review of the literature, Eur. J. Neurol. 7 (2000) 95–100.
[10] J.C. Horton, Arteriovenous malformations of the brain, N. Engl. J. Med. 357 (2007) 1774–1775.
[11] S. Inoue, W. Liu, K. Inoue, M. Youhei, K. Takenaka, H. Yamakawa, M. Abe, J.J. Jafar, R. Herzig, A. Koizumi, Combination of linkage and association studies for brain arteriovenous malformation, Stroke 38 (2007) 1368–1370.
[12] D.W. Johnson, J.N. Berg, M.A. Baldwin, C.J. Gallione, I. Marondel, S.-J. Yoon, T.T. Stenzel, M. Speer, M.A. Pericak-Vance, A. Diamond, A.E. Guttmacher, C.E. Jackson, L. Attisano, R. Kucherlapati, M.E.M. Porteous, D.A. Marchuk, Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2, Nat. Genet. 13 (1996) 189–195.
[13] S. Laberge-le Couteulx, H.H. Jung, P. Labauge, J.-P. Houtteville, C. Lescoat, C. Michaelle, E. Marechal, A. Joutel, J.-F. Bach, E. Tournier-Lasserve, Truncating mutations in CCM1, encoding KRIT1, cause hereditary cavernous angiomas, Nat. Genet. 23 (1999) 189–193.
[14] G.M. Lathrop, J.M. Lalouel, C. Julier, J. Ott, Strategies for multilocus linkage analysis in humans, Proc. Natl. Acad. Sci. U S A 81 (1984) 3443–3446.
[15] M. Lyons-Weiler, J. Hagenkord, C. Sciulli, R. Dhir, F.A. Monzon, Optimization of the Affymetrix GeneChip mapping 10K 2.0 Assay for routine clinical use on formalin-fixed paraffin-embedded tissues, Diagn. Mol. Pathol. 17 (2008) 3–13.
[16] N. Matsuzawa, K. Shimozato, N. Natsume, N. Niikawa, K. Yoshiura, A novel missense mutation in Van der Woude syndrome: usefulness of fingernail DNA for genetic analysis, J. Dent. Res. 85 (2006) 1143–1146.
[17] K.A. McAllister, K.M. Grogg, D.W. Johnson, C.J. Gallione, M.A. Baldwin, C.E. Jackson, E.A. Helmbold, D.S. Markel, W.C. McKinnon, J. Murrell, M.K. McCormick, M.A. Pericak-Vance, P. Heutink, B.A. Oostra, T. Haitjema, C.J.J. Westerman, M.E. Porteous, A.E. Guttmacher, M. Letarte, D.A. Marchuk, Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1, Nat. Genet. 8 (1994) 345–351.
[18] M. Nakashima, M. Nakano, A. Hirano, T. Kishino, S. Kondoh, N. Miwa, N. Niikawa, K. Yoshiura, Genome-wide linkage analysis and mutation analysis of hereditary congenital blepharoptosis in a Japanese family, J. Hum. Genet. 53 (2008) 34–41.
[19] A.M. Oberbauer, D.I. Grossman, M.L. Eggleston, D.N. Irion, A.L. Schaffer, N.C. Pedersen, J.M. Belanger, Alternatives to blood as a source of DNA for large-scale scanning studies of canine genome linkages, Vet. Res. Commun. 27 (2003) 27–38.
[20] N. Revencu, L.M. Boon, J.B. Mulliken, O. Enjolras, M.R. Cordisco, P.E. Burrows, P. Clapuyt, F. Hammer, J. Dubois, E. Baselga, F. Brancati, R. Carder, J.M.C. Quintal, B. Dallapiccola, G. Fischer, I.J. Frieden, M. Garzon, J. Harper, J. Johnson-Patel, C. Labreze, L. Martorell, H.J. Paltiel, A. Pohl, J. Prendiville, I. Quere, D.H. Siegel, E.M. Valente, A. van Hagen, L. van Hest, K.K. Vaux, A. Vicente, L. Weibel, D. Chitayat, M. Vikkula, Parkes Weber syndrome, vein of Galen aneurysmal malformation, and other fast-flow vascular anomalies are caused by RASA1 mutations, Hum. Mutat. 29 (2008) 959–965.
[21] N. Revencu, L.M. Boon, M. Vikkula, From germline towards somatic mutations in the pathophysiology of vascular anomalies, Hum. Mol. Genet. 18 (R1) (2009) R65–R74.
[22] A. Sasahara, H. Kasuya, H. Akagawa, H. Ujiie, O. Kubo, T. Sasaki, H. Onda, Y. Sakamoto, B. Krischek, T. Hori, I. Inoue, Increased expression of ephrin A1 in brain arteriovenous malformation: DNA microarray analysis, Neurosurg. Rev. 30 (2007) 299–305 discussion 305.
[23] K. Suphapeetiporn, P. Kongkam, J. Tantivatana, T. Sinthuwiwat, S. Tongkobpetch, V. Shotelersuk, PTEN c.511C>T nonsense mutation in a BRRS family disrupts a potential exonic splicing enhancer and causes exon skipping, Jpn. J. Clin. Oncol. 36 (2006) 814–821.
[24] U. Sure, E. Battenberg, A. Dempfle, W. Tirakotai, S. Bien, H. Bertalanffy, Hypoxia-inducible factor and vascular endothelial growth factor are expressed more frequently in embolized than in nonembolized cerebral arteriovenous malformations, Neurosurgery 55 (2004) 663–669.
[25] K. Takenaka, J. Muroi, S. Yamada, H. Yamakawa, M. Abe, K. Tabuchi, A. Koizumi, Genetic dissection of the familial cerebral arteriovenous malformation, Jpn. J. Neurosurg. 13 (2004) 837–845.
[26] R. Thiex, J.B. Mulliken, N. Revencu, L.M. Boon, P.E. Burrows, M. Cordisco, Y. Dwight, E.R. Smith, M. Vikkula, D.B. Orbach, A novel association between RASA1 mutations and spinal rteriovenous anomalies, AJNR. Am. J. Neuroradiol. 31 (2010) 775–779.
[27] J. van Beijnum, H.B. van der Worp, H.M. Schippers, O. van Nieuwenhuizen, L.J. Kappelle, G.J.E. Rinkel, J.W.B. van der Sprenkel, C.J.M. Klijn, Familial occurrence of brain arteriovenous malformations: a systematic review, J. Neurol. Neurosurg. Psychiatr. 78 (2007) 1213–1217.

Published online 6 May 2010

Nucleic Acids Research, 2010, Vol. 38, No. 17 ***5657–5671***
doi:10.1093/nar/gkq346

Genome-wide DNA methylation analysis in cohesin mutant human cell lines

Jinglan Liu[1], Zhe Zhang[2], Masashige Bando[3], Takehiko Itoh[3], Matthew A. Deardorff[1,4], Jennifer R. Li[1], Dinah Clark[1], Maninder Kaur[1], Kondo Tatsuro[5], Antonie D. Kline[6], Celia Chang[7], Hugo Vega[8], Laird G. Jackson[9], Nancy B. Spinner[1,4], Katsuhiko Shirahige[3] and Ian D. Krantz[1,4,*]

[1]Division of Human Genetics, Abramson Research Institute, [2]Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA, [3]Laboratory of Chromosome Structure and Function, Department of Biological Science, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, B2C 4259, Nagatsuta, Midori-ku, Yokohama City, Kanagawa 226-8501, Japan, [4]The University of Pennsylvania School of Medicine, PA 19104, USA, [5]Division of Developmental Disability, Misakaenosono Mutsumi Developmental, Medical, and Welfare Center, Konagai-cho Maki 570-1, Isahaya, 859-0169, Japan, [6]Harvey Institute for Human Genetics, Department of Pediatrics, Greater Baltimore Medical Center, Baltimore, MD 21204, [7]Genomic and Microarray Facility, the Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA, [8]Department of Genetics and Genomics Sciences, Mount Sinai School of Medicine, New York, NY 10029, USA and [9]Department of Obstetrics and Gynecology, Drexel University School of Medicine, Philadelphia, PA 19104, USA

Received January 1, 2010; Revised April 19, 2010; Accepted April 21, 2010

ABSTRACT

The cohesin complex has recently been shown to be a key regulator of eukaryotic gene expression, although the mechanisms by which it exerts its effects are poorly understood. We have undertaken a genome-wide analysis of DNA methylation in cohesin-deficient cell lines from probands with Cornelia de Lange syndrome (CdLS). Heterozygous mutations in *NIPBL*, *SMC1A* and *SMC3* genes account for ~65% of individuals with CdLS. SMC1A and SMC3 are subunits of the cohesin complex that controls sister chromatid cohesion, whereas NIPBL facilitates cohesin loading and unloading. We have examined the methylation status of 27 578 CpG dinucleotides in 72 CdLS and control samples. We have documented the DNA methylation pattern in human lymphoblastoid cell lines (LCLs) as well as identified specific differential DNA methylation in CdLS. Subgroups of CdLS probands and controls can be classified using selected CpG loci. The X chromosome was also found to have a unique DNA methylation pattern in CdLS. Cohesin preferentially binds to hypo-methylated DNA in control LCLs, whereas the differential DNA methylation alters cohesin binding in CdLS. Our results suggest that in addition to DNA methylation multiple mechanisms may be involved in transcriptional regulation in human cells and in the resultant gene misexpression in CdLS.

INTRODUCTION

Vertebrate gene expression is tightly regulated at several levels which are mechanistically linked to each other (1). DNA methylation, histone modification and chromatin remodeling are the most well-recognized and closely interwoven epigenetic events (2). Epigenetic regulation during development may occur very early during embryogenesis, driving the formation of different organ systems (3). DNA methylation is maintained by methyltransferase DNMT1 in both mitosis and meiosis, and is considered the most stable epigenetic mark (4).

In mammals, DNA methylation predominantly occurs at CpG dinucleotides by covalent addition of a methyl group to position 5 of the cytosine ring, creating 5-methylcytosine. CpG dinucleotides are underrepresented in the human genome with a frequency of 2–5% as compared to the GC content (5). CpG dinucleotides are not equally distributed throughout the human genome; instead, they occur in clusters of large repetitive

*To whom correspondence should be addressed. Tel: +1 215 590 2931; Fax: +1 215 590 3850; Email: ian2@mail.med.upenn.edu

sequences [such as ribisomal DNA (rDNA), satellite sequences or centromeric repeats] or in short CG-rich DNA stretches, known as CpG islands (CGIs) (6). Dinucleotide clusters of CpGs or 'CpG islands' are present in the promoter and exonic regions of ~40–70% of mammalian genes and these clusters are usually unmethylated (7,8). By contrast, other regions of the mammalian genome contain less CpG dinucleotides and the majority (75%) of these sparsely located CpG dinucleotides are largely methylated (9). A large number of experiments have shown that methylation of promoter CpG islands plays an important role in gene expression, genomic imprinting, X-chromosome inactivation, genomic instability, embryonic development and carcinogenesis (10,11).

Four DNA methyltransferases (DNMTs) (DNMT1, DNMT2, DNMT3A and DNMT3B) and one DNMT-related protein (DNMT3L) have been identified (12). DNMT1 acts as a maintenance methyltransferase, whereas DNMT2 may be an RNA methyltransferase, and DNMT3a and DNMT3b are *de novo* methyltransferases targeting unmethylated DNA. All DNMTs are essential for embryonic viability with homozygous mutant mice dying early in development (13). MBD1-4 proteins or methyl CpG-binding proteins (MeCP2) recognize and bind to methylated DNA. They recruit transcriptional corepressors such as histone-deacetylating complexes and polycomb group (PcG) proteins, associate with chromatin-remodeling complexes and attract chromodomain-binding proteins (13).

DNA methylation and chromatin structure are strikingly altered in many pathological situations, particularly in cancers and various mental retardation syndromes. Altered levels of the methyl donor folate and homocysteine have been repeatedly linked to these disorders. Disease-associated changes in epigenetic modifications can be classified into changes in genes that are epigenetically regulated and in genes that are part of the molecular machinery establishing and propagating the epigenetic modifications through the development and cell divisions. Aberrant methylation patterns have been reported in various neurodevelopmental disorders, including X-linked α-thalassemia and mental retardation (ATRX), Fragile X, and immune deficiency, centromeric instability and facial abnormalities (ICF) (14). Interestingly, the *ATRX* gene is misexpressed in CdLS [+1.2, false discovery rate (FDR) = 0.07] and the disorder ATRX presents defective sister chromatid cohesion and is considered as one of the 'cohesinopathies' (15).

Cornelia de Lange syndrome (CdLS, OMIM#122470, 300590, 610759) is the first identified 'cohesinopathy' which is a heterogeneous dominantly inherited developmental disorder with multiple-organ system involvement (16–19). The majority of CdLS probands were found to have heterozygous mutations in the *NIPBL* gene, whereas a small percentage have mutations in the *SMC1A* and *SMC3* genes. SMC1A and SMC3 are core components of the cohesin complex which controls sister chromatid cohesion during S phase, while NIPBL facilitates cohesin loading and unloading (20,21). In addition to cohesin and NIPBL's canonical role in regulating sister chromatid cohesion, they have also been implicated as key regulators of gene expression over long distances (22,23), and have been shown to preferentially associate to actively transcribed genes and colocalize with RNA polymerase II in *Drosophila* (24). In humans, cohesin colocalizes with CTCF and regulates gene expression at the imprinting *IGF2/H19* locus (25). Moreover, we recently performed a genome-wide transcription and cohesin-binding study on CdLS cell lines and identified a CdLS-specific expression profile from *NIPBL* and cohesin mutant individuals. Our data also suggest that cohesin preferentially binds to transcription start sites (TSSs) and tightly correlates to transcriptional activation in humans. Loss of cohesin binding occurs in CdLS and correlates to dysregulated gene expression (26). We undertook a genome-wide DNA methylation analysis on cell lines from probands with CdLS for the following reasons: First, the colocalization of CTCF and cohesin suggests possible functional overlap between these two proteins. A growing body of evidence suggests that CTCF is involved in regulating DNA methylation in vertebrates (14), and it binds to diverse DNA sequences including most imprinting center regions and many CpG islands. CTCF's binding to chromatin seems sensitive to DNA methylation, because it not only reads DNA-methylation marks but also has a role in determining DNA methylation patterns. Second, NIPBL was reported to be involved in chromatin remodeling by direct association with HP1, linking NIPBL to multiple enzymes and protein factors that determine the histone modification patterns that tightly correlate with DNA methylation status (27,28). Third, a qualitative analysis of cohesin binding peaks in control and CdLS cells using data obtained from our previous chromatin immunoprecipitation (ChIP)–chip assay (26) has revealed 6.85% of cohesin peaks overlap CpG islands, while 7.77% of the lost peaks in CdLS overlap CpG islands, hence CpG islands are overrepresented by 6.35% ($P = 0.0018$) among cohesin peaks that are lost in *NIPBL* mutant CdLS cells. Fourth, we have identified differentially expressed genes in CdLS from our previous genome-wide studies (using a relatively loosened cutoff of FDR < 0.2), multiple of which are critical in DNA methylation. Of these, there are several proteins directly involved in DNA methylation, such as the *de novo* DNA methyltransferase DNMT1 (−1.11, FDR = 0.15), the universal methyl-binding domain proteins MBD1 (−1.23, FDR = 0.02) and MeCP2 (+1.11, FDR = 0.06). In addition, PcG proteins which form the polycomb repressive complexes 1 and 2 (PRC1 and PRC2) are also dysregulated in CdLS. For example, CBX7 (+1.2, FDR = 0.13) and BMI1(+1.36, FDR = 0.02) in PRC1, EZH2 (−1.2, FDR = 0.05) and SUZ12 (−1.13, FDR = 0.199) in PRC2. Of note, Enhancer of Zeste homolog 2 (EZH2) was suggested to serve as a recruitment platform for DNA methyltransferases because it interacts with methyltransferases (DNMTs) and associates with DNMT activity *in vivo* (29). Another example is GADD45A (+1.54, FDR = 0.057) ectopic expression of which leads to the reduction in methylation at both specific gene loci and the total cellular 5-methylcytosine content (30). Fifth, SmcHD1-a SMC hinge domain

containing protein, maintains hypermethylation on the inactivated X chromosome in mice (31); and a second SMC-like protein-DMS3 mediates RNA-directed DNA methylation (RdDM) in plants (32), functional identification of these two proteins therefore links cohesin to DNA methylation. Sixth, 'genomic neighborhood diseases' have recently been proposed (33), and are composed of genes that are located in the same chromatin domain, are co-expressed and communicate through three dimensional structures. Dysregulated epigenetic events such as DNA methylation are tightly involved in gene mis-expression and the onset of these diseases. It is not known whether mutations in NIPBL or cohesin might be able to alter global DNA methylation and affect genomic organization, which may contribute to transcriptional dysregulation in CdLS.

Very little is known about epigenetic regulation by cohesin. Similarly, the role of epigenetic modulation in CdLS and the majority of other human developmental disorders are poorly characterized. To date, there has been no global assessment of overall DNA methylation in cohesin or *NIPBL* mutant human cells. We applied a comprehensive DNA methylation profiling approach to assess the epigenetic state in CdLS. We asked whether these mutant cells differed from healthy controls in terms of DNA methylation. We used an array-based method to quantitatively measure the methylation levels of 27 578 CpG sites in the regulatory regions of 14 495 genes in the human genome. We have identified differential methylation patterns in CdLS probands and also provided an integrative whole-genome view on DNA methylation, gene expression and cohesin binding in CdLS. We suggest CdLS has its own epigenetic signature that is formed in the early stage of embryonic development, which likely contributes to its clinical features.

MATERIALS AND METHODS

Sample collection

Sixty-three lymphoblastoid cell lines (LCLs) from 39 CdLS probands, two Roberts syndrome (RBS) probands and 22 gender- and race-matched healthy controls were tested. In addition, triplicates of one universally methylated DNA control (Zymo Research) and triplicates of each of two universally unmethylated DNA controls were also included (CHEMICON). All the tested CdLS probands have identified gene mutations and well-documented clinical features, including 22 severely affected probands with *NIPBL* mutations, eight mildly affected probands with *NIPBL* mutations, eight mildly affected probands with *SMC1A* mutations and one mildly affected proband with an *SMC3* mutation. These samples include most of the individuals studied in our previous gene expression project (26). All human subjects participating in this study were enrolled under an institutional review board-approved protocol of informed consent at The Children's Hospital of Philadelphia and Misakaenosono Mutsumi Developmental, Medical, and Welfare Center. All subjects were evaluated by one or more experienced clinicians. Gene mutations were confirmed by sequencing.

Cell culture and DNA isolation

LCLs were grown uniformly in RPMI 1640 with 20% fetal bovine serum (FBS), 100 U penicillin/ml, 100 μg streptomycin/ml sulfate and 1% L-glutamine as described previously. All 63 cell lines were grown anonymously and processed randomly. Genomic DNA was isolated from LCLs following the manufacturer's instruction (Gentra Systems). The ND-1000 (NanoDrop Technologies, Wilmington, DE, USA) were used to check DNA quality and quantity, respectively.

Bisulfite treatment of DNA samples

Prior to hybridization, bisulfite conversion of DNA samples was performed using the EZ DNA methylation kit (Zymo Research). Five-hundred nanograms DNA were used; a thermocycling program with a short denaturation step was included for bisulfite conversion (16 cycles of 95°C for 30 s followed by 50°C for 1 h).

Methylation assays by Infinium HumanMethylation27 BeadChip hybridization

After bisulfite treatment, each sample was whole-genome amplified (WGA) and enzymatically fragmented, and then applied to the BeadChips using Illumina-supplied reagents and conditions at Genomic Facility at the the Wistar Institute. HumanMethylation27 DNA Analysis BeadChip (Illumina), which carries 27 578 highly informative CpG sites derived from the well-annotated NCBI CCDS database (Genome Build 36) and spans more than 14 495 genes, was used for this experiment. Allele-specific primer annealing was followed by single-base extension using DNP- and Biotin-labeled ddNTPs. After extension, the array was fluorescently stained, scanned and the intensities of the unmethylated and methylated bead types were measured by a BeadArray Reader (34). Each methylation data point was represented by fluorescent signals from the M (methylated) and U (unmethylated) alleles and was recorded via a Methylation Module in BeadStudio software. DNA methylation values were described as beta (β) values which computed from the two alleles: $\beta = M/(U+M)$. The β-value therefore reflects the fractional methylation level of each CpG site. DNA methylation β-values are continuous variables between 0 and 1, representing the ratio of the intensity of the methylated bead type to the combined locus intensity. We quantified methylation level using β-value, and performed the statistic analysis based on the value of $\text{Log}_2[\beta/(1-\beta)]$ for more linearized data. We arbitrarily defined CpGs with $\text{Log}_2[\beta/(1-\beta)] > 0$ [also equals to $\beta > 0.5$ (50% of reads)] as hyper-methylated, CpGs with $\text{Log}_2[\beta/(1-\beta)] < -2$ [also equals to $\beta < 0.2$ (20% of reads)] as hypo-methylated, and CpGs with $-2 < \text{Log}_2[\beta/(1-\beta)] < 0$ as medium methylated.

Statistical analysis

Data processing and statistical analyses were preformed within R statistical environment (www.r-project.org); all CpG sites were included in the analysis. Array data from the methylated and unmethylated alleles were first processed separately by LOESS normalization across 63 LCL samples (the three sets of triplicated artificially methylated and unmethylated DNA controls were excluded) and then put together to calculate β-values. Differential methylation between control–CdLS or male–female groups was evaluated by a two-way analysis of variance (ANOVA) model with disease status and sample gender as the two tested factors. FDR was estimated by a procedure that randomly shuffled the sample labeling and repeated the ANOVA test 100 times.

Bisulfite sequencing validation

To validate the data obtained using the HumanMethylation27 BeadChip, samples from four to six healthy controls and four to six severely affected probands with *NIPBL* mutations were evaluated by bisulfite sequencing (BS). The genomic addresses and sequence information of each CpG dinucleotide on the BeadChip were downloaded from the company's database (Illumina); ±200 bp surrounding the target CpG of *CAPN2* and *LMO2* were retrieved from the UCSC genome database (http://genome.ucsc.edu/) and used as template sequence to design polymerase chain reaction (PCR) primers. Genomic DNA (1 µg) was bisulfite-converted and recovered as described above. Primers were designed by Methyl Primer Express v1.0 software (Applied Biosystems) using default settings. Primer sequences are available on request. Hot-start touchdown PCR was done in 25 µl reaction containing 0.25 mM dNTPs, 1× buffer, 0.4 µM forward and reverse primers and 1 U of Zymo*Taq*™ DNA Polymerase (Zymo Research). The PCR conditions were as the following: 94°C for 15 min, then 14 cycles with a gradual decrease of annealing temperature from 62°C to 55°C with 0.5°C reduction per cycle followed by 72°C for 1 min per cycle. After that, amplification was continued with 36 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min, then ended with 72°C for 15–20 min. PCR products were verified by gel electrophoresis, 2 ul PCR product was subsequently cloned into pGEM-T vector and transformed into JM109 cells according to the manufacturer's instruction (Promega). Ten to twelve individual clones of each PCR fragment were selected for sequencing using an ABI Prism 377 automatic sequencer (Applied Biosystems) with T7 primer. Sequencher and MacVector were used to align obtained sequences with reference sequences from the UCSC genome browser (http://genome.ucsc.edu/). C→T changes at CpG sites were documented.

Affymetrix expression array hybridization and ChIP microarray analysis

All procedures were performed as describe in our previous publication (26).

Online data mining

EpiGRAPH (http://epigraph.mpi-inf.mpg.de/WebGRAPH/) is an online software to analyze genomic and epigenomic features enriched in a group of given DNA fragments (35–37). EpiGRAPH was used to analyze DNA sequences harboring the differentially methylated CpG sites in CdLS in terms of the specific DNA sequence patterns, the overlap with specific genomic regions (e.g. CpG islands, repetitive regions and SNPs) and histone modification makers. Galaxy (http://galaxy.psu.edu/) was used to format downloaded genomic sequences from the UCSC genome browser (http://genome.ucsc.edu/). ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html) was used to identify consensus sequence around CpG sites whose methylation was significantly changed in CdLS.

Accession numbers

Genomic sequences reported in this manuscript have been submitted to NCBI GEO (http://www.ncbi.nlm.nih.gov/geo): methylation data are under accession number GSE 18458, gene expression data are under accession number GSE 12408 and ChIP–chip data are under accession number GSE 12603.

RESULTS

The overall whole-genome DNA methylation patterns in human LCLs in healthy controls

Whole-genome DNA methylation studies were conducted on 72 sodium bisulfite-converted DNA samples obtained from 63 sample LCLs, two sets of triplicated artificially de-methylated DNA controls and one set of triplicated artificially fully methylated DNA controls. Infinium HumanMethylation27 BeadChip that carries 27 578 CpG dinucleotides located within the promoter region of 14 495 unique genes in the human genome was used for hybridization. Hybridization signal and β-value were examined as described in 'Material and Methods' section, $\beta < 0.2$ (20% of reads) was considered as low level of methylation, while $\beta > 0.5$ (50% of reads) was considered as high level of methylation. Linearized $\text{Log}_2 [\beta/(1-\beta)]$ was used for statistic analysis. Samples were selected from Caucasian individuals and gender and age were closely matched. The 63 experimental samples included 22 healthy controls, 22 severely affected CdLS probands with *NIPBL* protein truncating mutations, eight mildly affected CdLS probands with *NIPBL* missense mutations, eight mildly affected CdLS probands with *SMC1A* mutations, one mildly affected proband with the *SMC3* mutation and two RBS probands with homozygous mutations in *ESCO2* (Supplementary Table S1). After normalization, the six replicated unmethylated and three replicated methylated artificial DNA control samples showed corresponding low and high β-values, indicating an efficient bisulfite conversion of DNA samples (Supplementary Figure S1). The methylation density peaks of all of the 63 experimental samples have $\text{Log}_2 [\beta/(1-\beta)] < 0$ and overlap the peaks of the unmethylated

artificial controls, indicating that the DNA from human LCLs is hypo-methylated globally (Supplementary Figure S1). There are 26 486 CpG dinucleotides on the human autosomes (chromosomes 1–22), the average DNA methylation levels of these dinucleotides are highly variable among the 22 control LCLs, but the majority of them are hypo-methylated. Bimodal distribution of DNA methylation was seen on all CpGs among the 22 controls, suggesting two types of genomic contents may be produced by DNA methylation. This bimodal distribution further suggests that it may represent the overall hypo-methylation of CpG sites inside CpG islands and the overall hyper-methylation of CpG sites located within the CpG poor genomic regions in human LCLs (Supplementary Figure S1).

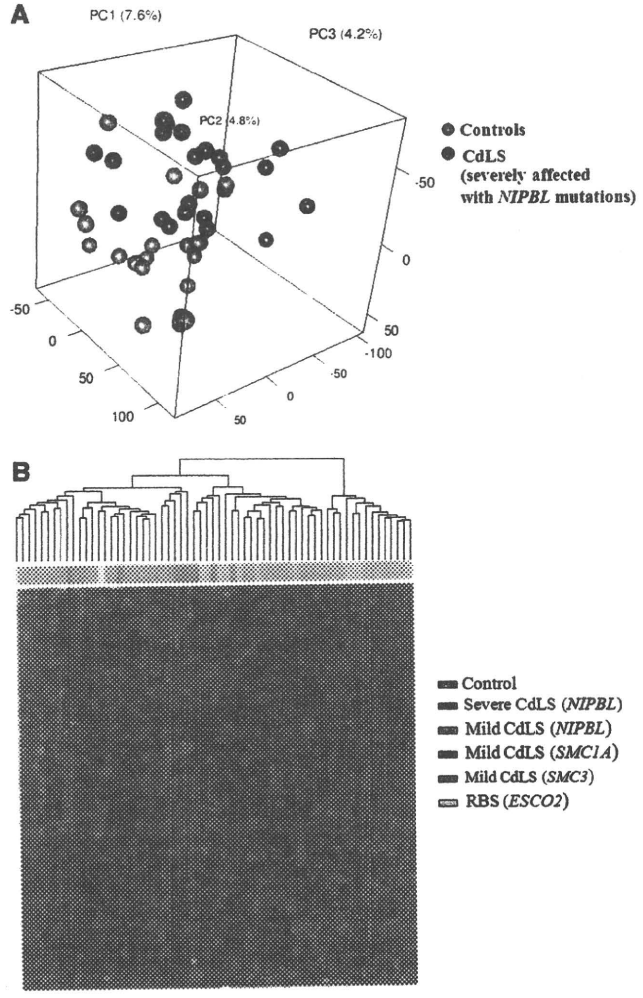

Figure 1. CdLS specific DNA methylation profile revealed by unsupervised and supervised classification analysis. (**A**) Unsupervised principal component analysis (PCA) of DNA methylation on 26 486 autosomal CpG sites. Twenty-two severely affected CdLS probands with *NIPBL* mutations and 22 healthy controls are clearly separated into two groups except five controls. $Log_2[\beta/(1-\beta)]$ was calculated for each probe in each sample. (**B**) Hierarchial clustering of 63 experimental samples based on 152 differentially methylated autosomal CpG sites ($P < 0.001$). Individuals with different clinical presentations are color-coded, the corresponding gene mutations are labeled inside the brackets.

Identification of CdLS specific DNA methylation patterns in LCLs

Differential DNA methylation identified in CdLS probands. Methylation levels of 26 486 probes on autosomes (chromosomes 1–22) were quantified by β, afterwards $Log_2[\beta/(1-\beta)]$ of each probe in each sample was calculated and used for unsupervised principal component analysis (PCA). Figure 1A depicts that the 22 severely affected CdLS probands with *NIPBL* protein truncating mutations can be separated from the 22 healthy controls solely based on DNA methylation levels on autosomes except five healthy control samples. Control 'AGS-222-S', '27574', 'CDL-145-03S', '27572-B' and '95-3986-S' are discordant to the rest of other samples in control group. This unsupervised PCA result indicates that the genome-wide DNA methylation profile is significantly different in CdLS as compared to healthy individuals. Probes on X and Y were excluded due to the gender effect on DNA methylation.

In order to define the differentially methylated DNA loci in CdLS, we performed a two-way ANOVA analysis of the total 27 578 probes on the control group (22 individuals) and the severely affected *NIPBL* mutant CdLS group (22 individuals). There are 924 CpG sites (corresponding to 902 cognate genes) differentially methylated in CdLS with $P < 0.01$ (FDR = 0.222), out of these 924 CpG sites, methylation levels were decreased on 361 sites (356 genes) and increased on 563 (546 genes) sites in CdLS (Supplementary Table S2). We selected 152 differentially methylated CpG sites on the autosomes with $P < 0.001$ (Supplementary Table S2) and further performed clustering analysis on all of the 63 samples in our cohort which are varied both clinically and genotypically (Figure 1B). Samples from each subgroup could be clustered together, although with a few outliers. Control samples and severely affected *NIPBL* mutant CdLS probands are evidently separated from each other, whereas mildly affected CdLS probands stay in between. Within the mildly affected CdLS probands, individuals with *NIPBL* mutations stay closer to the *NIPBL* mutant subgroup that has severe manifestations, whereas CdLS probands with other gene mutations (*SMC1A* and *SMC3*) tend to cluster together and stay closer to healthy controls. Interestingly, the two RBS probands are clustered side by side with *SMC1A* mutant individuals. RBS is an autosomal recessive genetic disorder due to homozygous or compound heterozygous mutations in the *ESCO2* gene, ESCO2 has acetyltransferase activity and is involved in the establishment of sister chromatid cohesion.

The X chromosome has a unique DNA methylation pattern in CdLS. We further separately analyzed the hypo-methylated CpG sites ($Log_2[\beta/(1-\beta)] < -2$) and hyper-methylated CpG sites ($Log_2[\beta/(1-\beta)] > 0$) in controls. In addition, autosomal and X-linked sites were also analyzed separately (Supplementary Table S3). In general, the average DNA methylation levels of all of the hyper-methylated CpG sites and the X-linked hypo-methylated CpG sites have been increased in

CdLS, whereas the autosomal hypo-methylated CpG sites remain unchanged (Figure 2A). Statistically, the average methylation level (combining both females and males) of X-linked hypo-methylated CpG sites has increased significantly in CdLS as compared to the increase of autosomal hypo-methylated CpG sites ($P = 1.51E-08$) (Figure 2A). Since gender could play a major role affecting DNA methylation levels, we then repeated the above analysis by analyzing female and male samples separately. We saw more significant methylation change of hypo-methylated CpG sites on the X chromosome than on the autosomes in CdLS females ($P = 3.5E-17$). Significant increase in DNA methylation of hypo-methylated CpG sites than that of hyper-methylated CpG sites was also found on X chromosome in female CdLS probands ($P = 1.2E-12$) (Figure 2B). Surprisingly, significantly reduced methylation levels were found for the above hypo-methylated CpG sites on the X chromosomes in male CdLS probands (Figure 2B).

The correlation between the differential DNA methylation and the dysregulated gene expression in CdLS

General correlation between DNA methylation and gene expression in control human LCLs. The expression of genes in LCLs and the methylation levels of their associated CpG sites were first examined in 22 control samples. We took advantage of the whole-genome expression analysis from our previous study (26), in which we have identified 10 378 out of 15 162 unique refSeq genes expressed in LCLs using Affymetrix HG_U133plus2.0 arrays. Each of these genes was mapped to one and only one transcription starting site (TSS). CpG dinucleotides were mapped to within ± 1-kb regions surrounding TSSs to obtain 22 351 TSS–CpG pairs from 12 081 unique genes and 21 110 unique CpG sites. Consistent with the well-documented literature, an inverse correlation between DNA methylation and gene expression was also found in LCLs in our study. DNA methylation suppresses gene expression and the reduced methylation level correlates to higher probability of gene expression (Figure 3A, Supplementary Figure S2A). Figure 3A shows the percentage of expressed genes and the corresponding CpG methylation levels. The overall percentage of genes expressed in LCLs is 69.9% (8449/12 081), which indicates that 69.9% of human genes are likely to be expressed in LCLs. While the level of DNA methylation is increased, fewer genes are expressed. Using ($\text{Log}_2\ [\beta/(1 - \beta)] = 0$) as cutoff, Fisher's exact test shows highly significant interdependence between gene expression and DNA methylation in the promoter region ($P = 4.5E-315$) (Figure 3A). The elevated expression of a subgroup of genes that is associated with hyper-methylation ($\text{Log}_2[\beta/(1 - \beta)] > 0$) seems to be unexpected, suggesting mechanisms other than DNA methylation are involved in their transcriptional regulation; or in another word, DNA methylation alone is not enough to downregulate gene expression in human LCLs (Figure 3A, Supplementary Figure S2A). We subdivided the 21 110 unique sites into two groups based on the CpG island mapping information from the UCSC genome browser (http://genome.ucsc.edu) as CpG sites located within the CpG island (CGI sites) and outside the CpG island (non-CGI sites). There was no linear correlation between the level of DNA methylation and the transcriptional activity of the cognate gene observed for both CGI sites and non-CGI sites, which further indicated the direct impact from DNA methylation on gene expression might not be predominant (Supplementary Figure S2B).

We also found that the relative location of the CpG site to the TSS impacts the association between DNA methylation and the expression of the downstream gene (Figure 3B). While 44.5% (1224/2752) of the genes with hyper-methylated CpG sites around their TSSs were expressed in LCLs, the closer the sites were located to the TSSs, the less likely the genes were to be expressed. When there was a hyper-methylated site located within ± 50 bases around a TSS, the downstream gene had only a 27.6% (93/337) chance of being expressed. However, such an effect of the hypo-methylated sites was less evident. The negative correlation between DNA methylation and gene expression is stronger on autosomes than on the X chromosome (Supplementary Figure S2C). The probability of genes to be expressed on autosomes and the X chromosome is 82.3% and 61.4%, respectively ($P = 6.0E-06$) when there is a hypo-methylated site around the TSS in female samples, and 36.8% and 44.5% ($P = 0.08$) when there is a hyper-methylated site. This observation suggests that the association between DNA methylation and gene expression has different mechanisms on autosomes and the X chromosome.

The correlation between differential DNA methylation and altered gene expression in CdLS. As described earlier, 12 081 unique refSeq genes are associated with CpG sites with DNA methylation information available from this study. Out of these 12 801 genes, 8449 genes are expressed in LCLs. We compared the expression level and the methylation level for each of these 8449 genes between 22 controls and 22 severely affected *NIPBL* mutant CdLS probands, then we plotted the alteration of expression to the alteration of methylation for each gene. In general, there is no strong evidence supporting that global differential DNA methylation correlates to global transcriptional dysregulation in CdLS (Figure 4). The increased DNA methylation level correlates to the transcription downregulation for only six genes in CdLS ($P = 0.003$), whereas the decreased DNA methylation level correlates to transcriptional upregualtion for only two genes but without statistic significance (Figure 4, Supplementary Table S4). There is no link seen between the altered gene expression of the rest of 8441 genes and the altered DNA methylation in CdLS (Figure 4). We therefore conclude that the differential DNA methylation may not directly contribute to the transcriptional dysregulation in CdLS.

The correlation between the differential DNA methylation and cohesin binding in CdLS

Cohesin preferentially binds to hypomethylated DNA in control human LCLs. The average DNA methylation

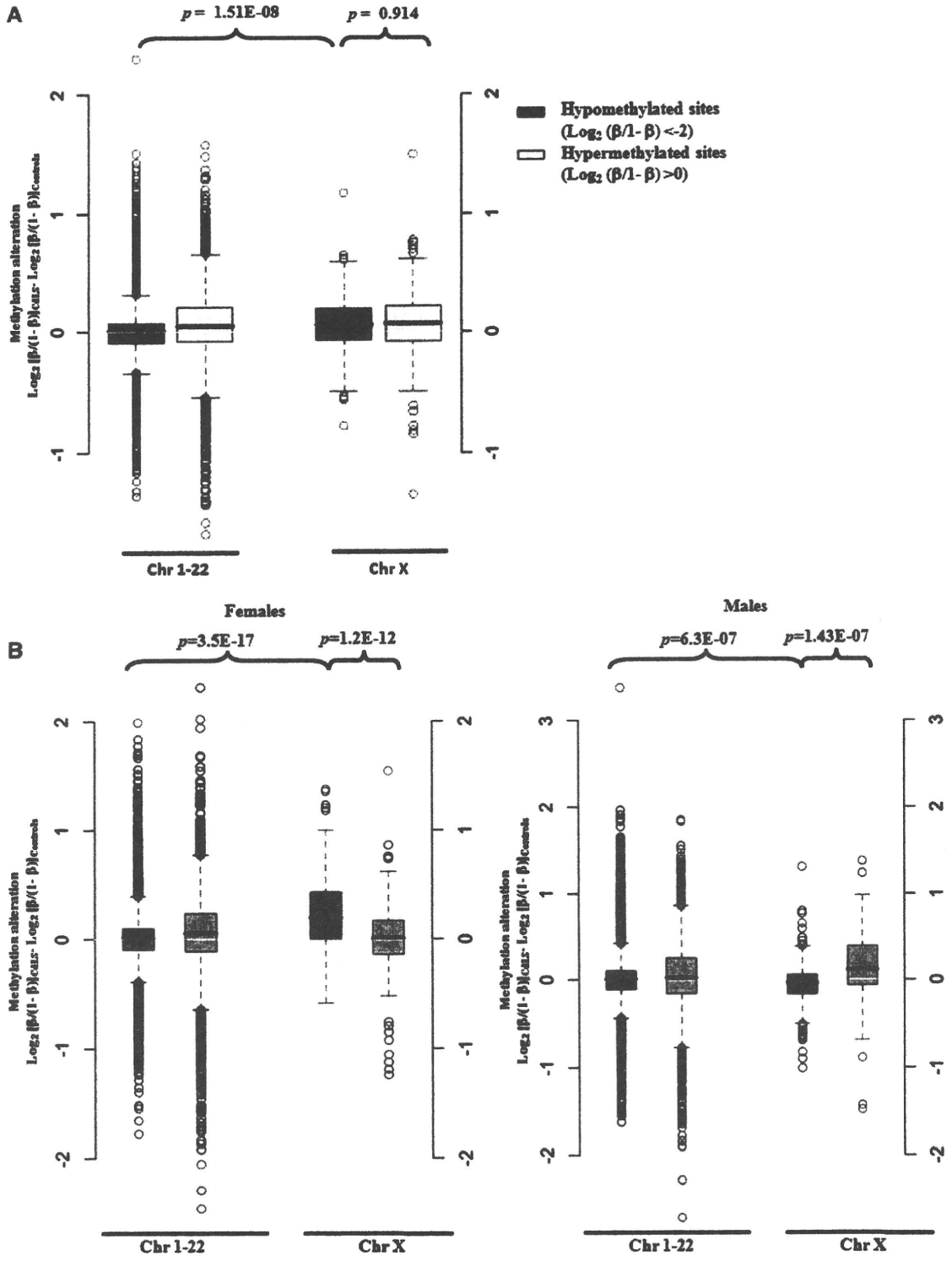

Figure 2. Box plots showing DNA methylation changes for hypo-methylated CpG sites on chromosome X are different in CdLS. (**A**) DNA methylation levels of X-linked hypo-methylated CpG sites have significantly increased in CdLS as compared to autosomal hypo-methylated CpG sites. Twenty-two controls and 22 severely affected CdLS probands with *NIPBL* mutations were included. (**B**) DNA methylation levels of X-linked hypo-methylated CpG sites have significantly increased in female CdLS probands but decreased in male CdLS probands as compared with methylation changes of the autosomal CpG sites. Twenty-two controls and 22 severely affected CdLS probands with *NIPBL* mutations were included, and females (11 controls and 11 probands) and males (11 controls and 11 probands) were analyzed separately. The label of *Y*-axis is defined as 'CdLS-Control' = $\text{Log}_2\ (\beta/1-\beta)_{\text{CdLS}} - \text{Log}_2\ (\beta/1-\beta)_{\text{Control}}$, indicating the alteration of methylation levels between CdLS and controls.

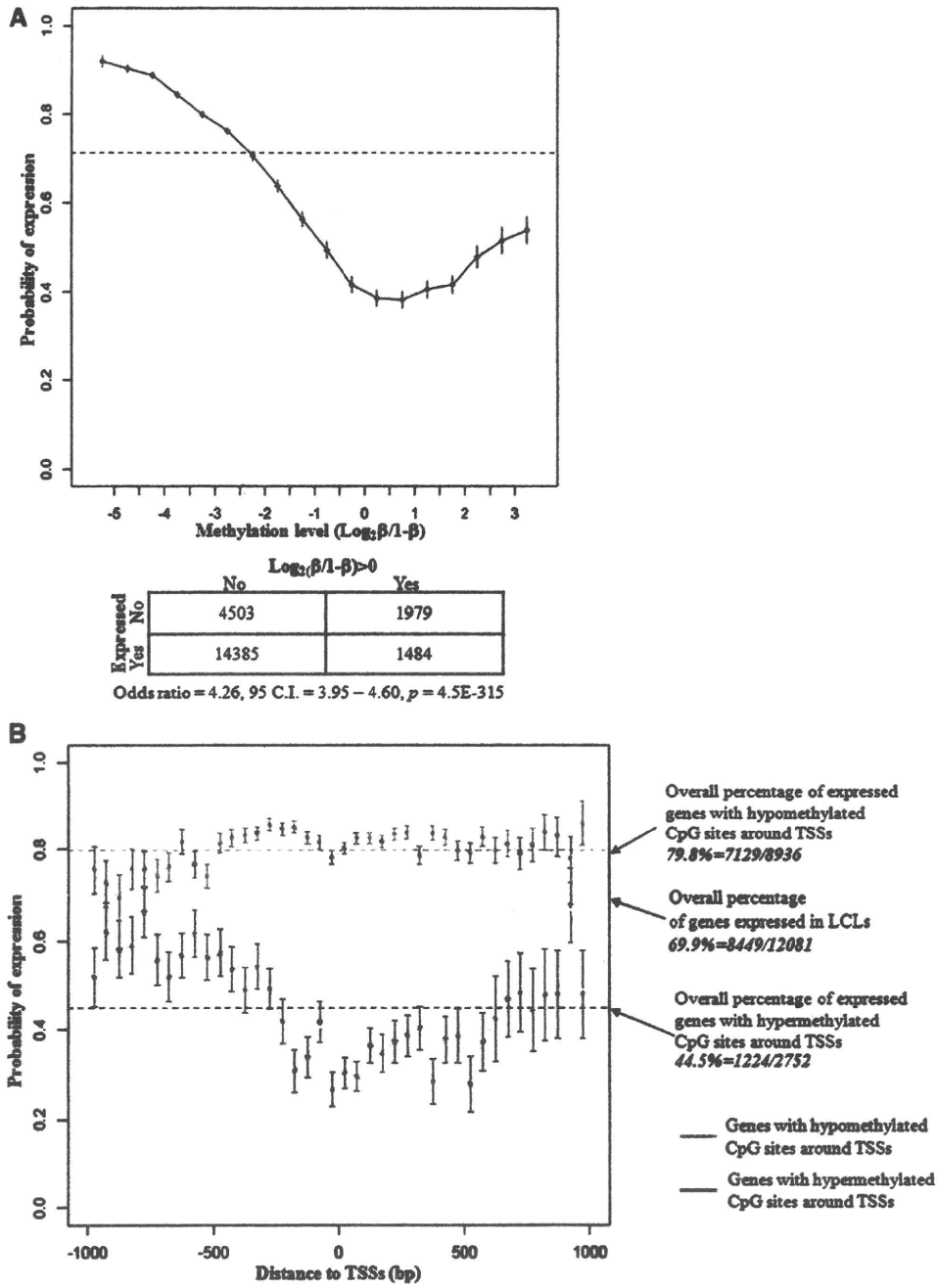

Figure 3. Correlation between the DNA methylation and the gene expression in human LCLs. (**A**) Increased level of DNA methylation correlates to reduced gene expression. A total of 22 351 TSS–CpG pairs from 12 081 unique genes were analyzed. The broken line indicates that 8449 (69.9%) of these genes were expressed in LCLs. (**B**) Location of the CpG sites impacts the association between DNA methylation and gene expression. Among the genes having hyper-methylated CpG sites around their TSSs, the closer the CpG sites are located to TSS, the less possible the downstream gene is expressed (red lines). This effect is not obvious among genes with hypo-methylated CpG sites around their TSSs (green lines).

level of 21 110 CpG sites associated to the above 12 081 genes in 22 controls were analyzed with $\text{Log}_2[\beta/(1-\beta)]$. Cohesin binding signal was obtained from our previous ChIP–chip assay (26), re-analyzed quantitatively and mapped to each of 21 110 CpG site. The majority of cohesin signals were concentrated at chromatin regions with a low level of DNA methylation which is $\text{Log}_2[\beta/(1-\beta)] < -2$ ($\beta < 0.2$), indicating that cohesin preferentially binds to hypo-methylated chromatin; meanwhile, an increased DNA methylation level prohibits cohesin from binding to the chromatin (Supplementary Figure S3A and B). From our previous study (26), we have shown evidence that cohesin preferentially binds to gene promoters especially at TSSs. We therefore split the 12 081 genes into three groups according to the methylation levels of the CpG sites at their TSSs, and looked at cohesin binding intensities around promoters. As described above, CpGs with $\text{Log}_2[\beta/(1-\beta)] > 0$ were

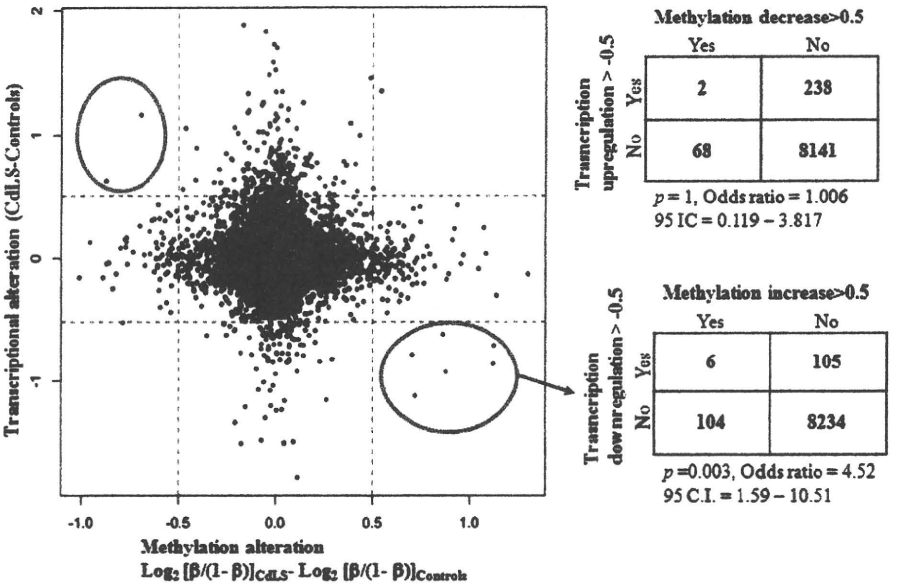

Figure 4. The change of DNA methylation does not generally affect gene expression in CdLS. The 8449 genes that were expressed in LCLs were examined in 22 controls and 22 severely affected probands. The increased methylation shows a significant suppression effect on transcription for only six genes in CdLS with $P = 0.003$ in Fisher's test. The decreased DNA methylation reveals no significant correlation with transcriptional upregulation in CdLS. *Y*-axis represents the gene expression changes in CdLS calculated with signal intensity on expression arrays; *x*-axis represents the methylation alteration in CdLS using $\text{Log}_2\ (\beta/1 - \beta)$. The 0.5 cutoff value was set up arbitrarily.

defined as hyper-methylated, while CpGs with $\text{Log}_2[\beta/(1-\beta)] < -2$ were defined as hypo-methylated, and CpGs with $-2 < \text{Log}_2[\beta/(1-\beta)] < 0$ were defined as medium methylated. A similar cohesin-binding pattern was identified for the 8128 genes with hypo-methylated CpGs and the 2247 medium methylated CpGs. Cohesin preferentially binds to the vicinity of their TSSs especially around the core promoter region (from −200 bp to +1 bp) with each peak clearly seen (Figure 5). On the contrary, very little cohesin binds to the 1706 genes with hyper-methylated CpGs, and there is no quantitative difference of cohesin binding to different regions surrounding the TSSs (Figure 5). Although cohesin binding correlates to gene expression (26), we were not able to identify a direct correlation between gene expression and DNA methylation based on our above analyses. In combination with our previous studies, these results suggest the association between cohesin and promoters are tightly correlated to gene expression, but DNA methylation may only be one of several upstream events that affect cohesin binding. The transcriptional regulation does not appear to solely depend on DNA methylation but rather multiple mechanisms or pathways are likely functioning together to control the expression of genes in human LCLs.

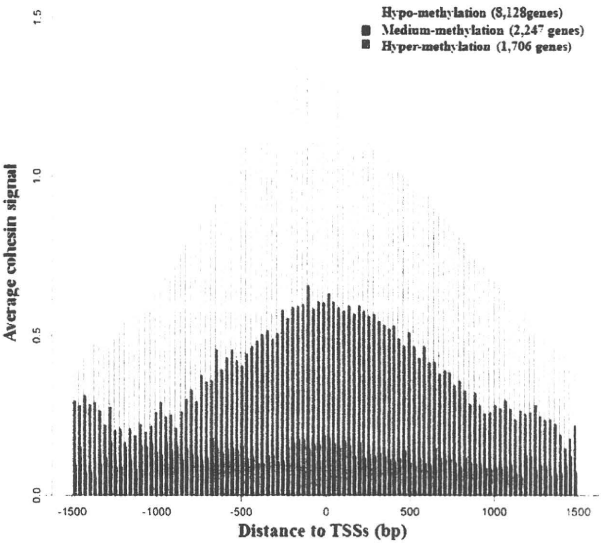

Figure 5. Cohesin preferentially binds to CpG sites at promoter regions with low levels of DNA methylation. Very little cohesin binds to hyper-methylated promoters regardless of the distance to TSSs. $\text{Log}_2(\beta/1-\beta) > 0$ was defined as hyper-methylation, $\text{Log}_2(\beta/1-\beta) < -2$ was defined as hypo-methylation, $-2 < \text{Log}_2(\beta/1-\beta) < 0$ was defined as medium methylation. Cohesin binding signal was quantified from previous ChIP–chip study and log-transformed before the analysis.

Differential DNA methylation alters cohesin binding in CdLS. In addition to respectively examining genes with hypo-methylated promoters and hyper-methylated promoters in CdLS, we also wanted to examine whether the X chromosome or autosome location of the CpG dinucleotide will differently affect the correlation between DNA methylation and cohesin binding in CdLS, hence only the female samples from the control group and severely affected *NIPBL* mutant CdLS individuals were analyzed. For the genes with hypo-methylated promoters, binding of cohesin is enriched around TSSs regardless of whether they are located on autosomes (6692 genes) or X chromosome (145 genes); however,

less cohesin binds to the X chromosome than to autosomes and the peak shifts upstream of the TSS. In CdLS, the amount of both autosomal and X chromosome bound cohesin has decreased, although a smaller peak remains in a narrowed region surrounding the TSS (Figure 6A). For the genes with hyper-methylated promoters, the amount of cohesin bound is quite low for both autosomal and X-linked genes (1640 and 129, respectively) (Figure 6B), and cohesin binding is equally distributed along ±1.5-kb region of TSSs with no obvious peak seen at TSSs. More cohesin is associated with the X chromosome than with the autosomes in controls but this tendency was reversed in CdLS which shows less cohesin associated with the X chromosome than with autosomes (Figure 6B). A slightly reduced amount of cohesin binds to the X chromosome in CdLS, but there is little change in cohesin binding to autosomal genes in CdLS (Figure 6B). In summary, cohesin binding has changed remarkably at hypo-methylated promoters in CdLS, with unknown mechanisms in addition to DNA methylation likely involved in regulating cohesin binding to X-linked and hyper-methylated promoters.

BS validation for microarray results and evidence of varied DNA methylation in control LCLs

Bisulfite conversion and sequencing described by Susan Clark (38) has always been considered the gold standard by which to measure CpG methylation. We selected two genes, *LMO2* and *CAPN2*, which demonstrated differential methylation between CdLS and control cells, for BS validation. In our previous studies (26), these two

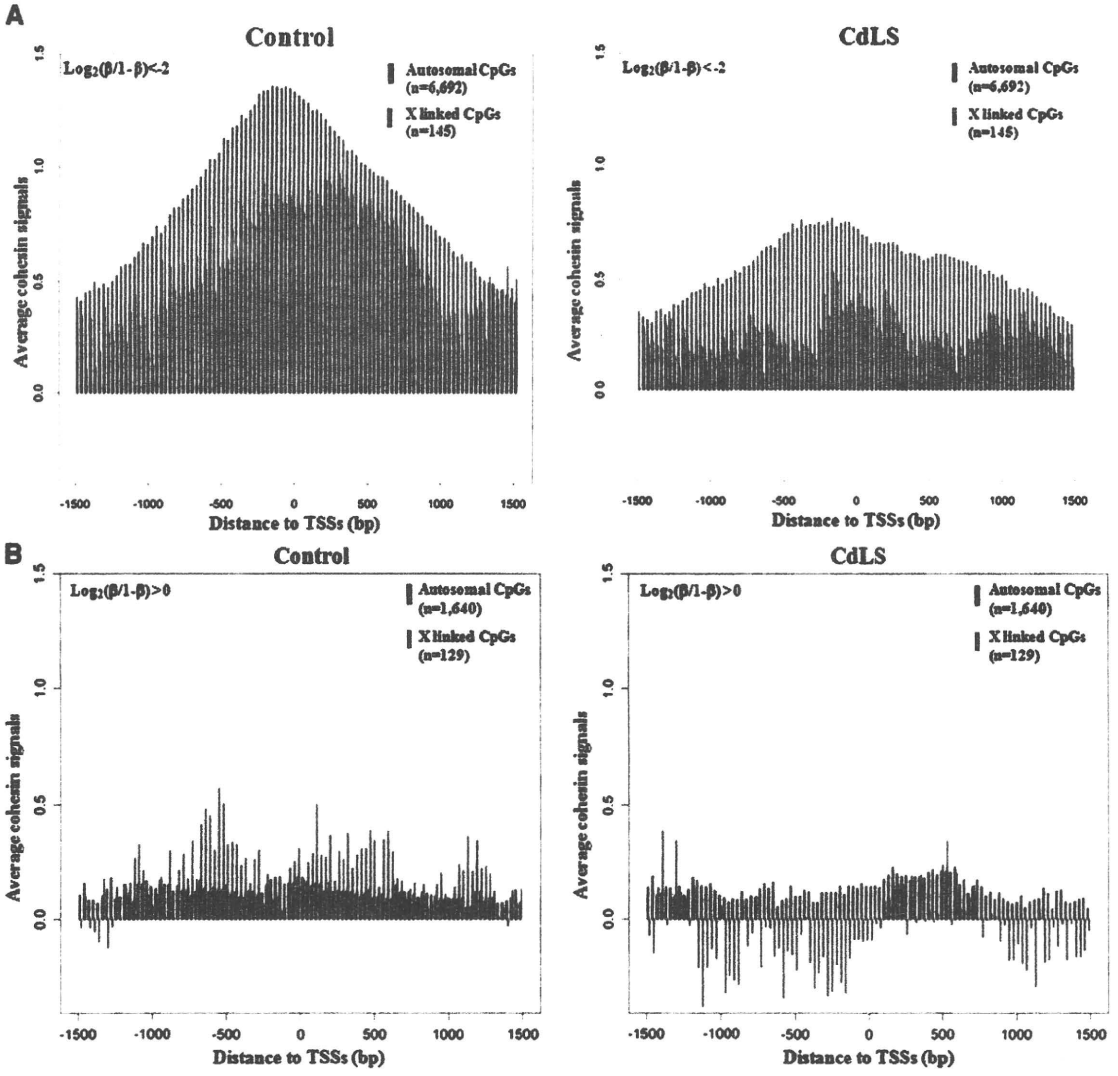

Figure 6. Reduced cohesin binding at hypo-methylated promoters in CdLS with minimal changes for hyper-methylated promoters. (**A**) Cohesin binding to genes with hypo-methyated promoters is reduced in CdLS, and less cohesin binds to promoters on the X chromosome than on autosomes; 6692 of such genes are on autosomes, while 145 are on the X chromosome. (**B**) Less cohesin binds to genes with hyper-methylated promoters in both controls and probands; however, more cohesin binds to X-linked promoters than to autosomal promoters in controls and this X-linked cohesin binding is reduced in CdLS. $\text{Log}_2(\beta/1-\beta) < -2$ was defined as hypo-methylation and $\text{Log}_2(\beta/1-\beta) > 0$ was defined as hyper-methylation. Cohesin signal was quantified and log-transformed before the analysis. Only 11 female samples from controls and 11 female samples from severely affected CdLS probands were analyzed.