Cytogenet Genome Res 123:169–175 (2008)
DOI: 10.1159/000184705

Cytogenetic and Genome Research

Copy number variation in metabolic phenotypes

M. Lanktree R.A. Hegele

Vascular Biology Group, Robarts Research Institute, and Departments of Medicine & Biochemistry, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ont. (Canada)

Accepted in revised form for publication by H. Kehrer-Sawatzki and D.N. Cooper, 16 June 2008.

Abstract. Despite successes in identifying genetic contributors to common metabolic phenotypes, only part of the heritable component of these traits has thus far been explained. Copy number variation (CNV) is likely to be responsible for some of the unexplained variation. As observed with single nucleotide changes, it is probable that both rare and common CNVs will contribute to susceptibility to metabolic disease. For instance, CNVs in the *LDLR* gene underlie a substantial portion of disease in patients with heterozygous familial hypercholesterolemia. As well, a common CNV in *LPA* encoding apolipoprotein(a) is the primary determinant of plasma lipoprotein(a) concentrations, a risk factor for atherosclerosis. Recent efforts to map CNVs in control populations have defined their size, frequency and distribution. Many of the identified CNVs overlap genes with important functions in metabolic pathways. The overlap of CNVs that were found in control datasets with functional candidate genes or genes with previous evidence of association with metabolic syndrome presents an important subset for future CNV association studies. Finally, we describe an approach to search for CNVs in a rare high-penetrance metabolic disorder, namely lipodystrophy. As methods to identify CNVs increase in precision and accuracy, the prospect of identifying their role in both rare Mendelian and common complex metabolic phenotypes will become a reality.
Copyright © 2009 S. Karger AG, Basel

Genetic investigations have successfully identified single nucleotide variants contributing to metabolic phenotypes generally following two models, namely, the ‘common variant-common disease’ model (Reich and Lander, 2001) and the ‘heterogeneity’ model (Wang WY et al., 2005). In the former, common genetic variants produce small-effect changes to protein structure, function or expression leading to small increases or decreases in disease susceptibility or in levels of a quantitative trait. For a deleterious variant to become common in a population, the negative effect of the variant must be sufficiently small for it to persist despite selection. Thus, common variants are usually of the small-effect type. In contrast, in the ‘heterogeneity’ model of complex disease, many rare mutations underlie susceptibility to disease, which is often defined in individuals who are clustered at the extreme of a quantitative trait. Rare mutations often have large effects, markedly affecting protein expression or function, leading to overt phenotypes that follow Mendelian inheritance patterns, such as low-density lipoprotein receptor *(LDLR)* mutations in heterozygous familial hypercholesterolemia (HeFH) or ATP-binding cassette transporter A1 *(ABCA1)* mutations in Tangier disease (Hegele, 2001). Genetic heterogeneity, in which variants in a range of different genes all lead to a similar phenotype, can be subsumed within either model of disease. Linkage disequilibrium between variants complicates matters, but if many individually rare mutations are found on the same ancestral haplotype, a surrogate common variant can still be used as a marker of disease susceptibility.

Deep resequencing methods attempt to identify the common and rare single nucleotide changes and small inser-

This work was supported by operating grants from the Heart and Stroke Foundation of Ontario (NA 6018), the Canadian Institutes for Health Research (MOP 13430 and 79533), by the Jean Davignon Distinguished Cardiovascular-Metabolic Research Award (Pfizer, Canada) and Genome Canada through the Ontario Genomics Institute. M.L. is supported by the Canadian Institute of Health Research MD/PhD Studentship Award, and the University of Western Ontario MD/PhD Program.

Request reprints from Robert A. Hegele MD FRCPC FACP
Blackburn Cardiovascular Genetics Laboratory
Robarts Research Institute, London, Ont., N6A 5K8 (Canada)
telephone: +1 519 663 3461; fax: +1 519 663 3037; e-mail: hegele@robarts.ca

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2009 S. Karger AG, Basel
1424–8581/08/1234–0169$24.50/0

Accessible online at:
www.karger.com/cgr

tions or deletions affecting disease susceptibility by sequencing relatively small regions of the genome in many individuals, typically beginning with those who are clustered at the extremes of a quantitative phenotype (Ahituv et al., 2007). Next generation DNA sequencers will provide rare variant information on a much larger scale and will stress bioinformatic capacity to separate variants that impact upon susceptibility from those that are neutral. Microarrays are also now commonly used to test common variants across the whole genome in order to find associations with disease susceptibility. The aggregate of evidence to date suggests that both common and rare variants contribute to overall susceptibility to metabolic disease phenotypes, under a hybrid or combination of the 'common variant-common disease' and 'heterogeneity' models.

Despite successes in identifying genetic contributors to metabolic phenotypes, unexplained heritability exists and the development of new strategies and the examination of new types of variation may help reduce the unexplained proportion. Improved precision and accuracy in identifying copy number variations (CNVs) will allow for these variants to be tested for their association with susceptibility to common diseases. CNVs are defined as quantitative increases or decreases in copy number (duplications or deletions) above the resolution of sequencing, but below the resolution of microscopic evaluation (>1 kb), while the term insertion/deletion (indel) is referring to copy number changes from 1 bp to 1 kb (Scherer et al., 2007). The terms CNV and indel do not imply anything about the frequency or phenotypic effect of the variation (Scherer et al., 2007). As with other types of genetic variation, it is likely that both common CNVs with small effect sizes and rare CNVs with large effect sizes will each play a role in complex diseases. Thus, studies should be designed with methods tailored to detect CNVs of different frequencies. In this paper, we review two types of CNVs that are associated with metabolic phenotypes: 1) rare CNVs that are causative for HeFH; and 2) a common CNV resulting in altered plasma concentrations of lipoprotein(a) (Lp(a)). In addition, we demonstrate how CNVs observed in control populations overlap genes that are involved in susceptibility to the metabolic syndrome. Finally, we report a preliminary approach to screen for a rare large-effect CNV in a rare metabolic disorder, namely lipodystrophy.

Rare CNVs causing HeFH

HeFH (MIM 143890) is an autosomal dominant disorder affecting ~1 in 500 people, characterized by the failure of the liver to catabolize plasma low-density lipoprotein (LDL) cholesterol (Yuan et al., 2006). The resulting elevation of plasma LDL cholesterol concentrations is associated with development of cholesterol deposits in the cornea, called corneal arcus, on their eyelids and extensor tendons, called xanthelasmas and xanthomas, respectively, and in their vessel walls, leading to atherosclerosis and up to 100 times greater risk of coronary heart disease (CHD) for men under 45 years of age (Yuan et al., 2006; Pollex and Hegele, 2007). The molecular mechanism for disease in the majority of HeFH patients are small insertions or deletions, or missense, splicing, or nonsense mutations within the LDL receptor gene (*LDLR*, MIM 606945), the majority of which were identified through sequencing of the coding exons and intron-exon boundaries using an approach called exon-by-exon sequence analysis (EBESA) (Wang et al., 2001). Over 500 different missense mutations and over 200 small insertions and deletions have been reported within *LDLR* (Stenson et al., 2003). Genetic heterogeneity is indicated by phenocopies of HeFH resulting from dominant mutations in genes encoding apolipoprotein (apo) B-100 (*APOB*, MIM 107730) and proprotein convertase subtilisin/kexin 9 (*PCSK9,* MIM 607786). Despite efforts to identify additional variants, approximately one third of patients clinically diagnosed with HeFH in some series have had no mutations identified (Pollex and Hegele, 2007).

The first large deletion reported in *LDLR* was described in 1985 when reduced LDLR protein size was observed in cultured fibroblasts from an HeFH patient (Lehrman et al., 1985). Southern analysis revealed a 5-kb deletion, including two and a half exons, resulting from uneven recombination between repetitive *Alu* elements (Lehrman et al., 1985). The same group later identified an HeFH patient with a duplication of seven exons in *LDLR* (Lehrman et al., 1987). Since then, >100 large deletions and duplications have been reported in HeFH patients, most of which were discovered with long range PCR or Southern analysis (Stenson et al., 2003). In 2004, the first paper studying prevalence of CNVs in *LDLR* using multiplex ligation-dependent probe amplification (MLPA) found that of 70 unrelated HeFH patients in southern Ontario, Canada, 44 had *LDLR* mutations and five had *APOB* mutations identifiable by EBESA, while 12 patients had deletions in *LDLR*, suggesting CNVs were responsible for ~17% of HeFH cases (Wang J et al., 2005). In the second paper examining *LDLR* with MLPA, 44 of 431 (10.1%) of HeFH patients from Norway possessed a CNV (Holla et al., 2005). A third paper reported only 5 of 318 (1.5%) HeFH patients had an *LDLR* CNV in a Danish study (Damgaard et al., 2005). Finally, a British study found 13 of 153 (8.5%) HeFH patients had an *LDLR* CNV (Tosi et al., 2007). MLPA was reported to be more time-efficient, and importantly more sensitive, than Southern analysis (Damgaard et al., 2005; Holla et al., 2005). Possible explanations for the variation in observed CNV frequency among HeFH patients includes differences in assay efficiency between studies, although all studies used MLPA kits from MRC-Holland (Amsterdam, the Netherlands), differences in study populations, although most patients were European Caucasian in origin, differences in the founder populations of the geographical locations studied, or sampling error. Pooling of the results indicates that 74/972 (7.6%) of HeFH patients have a CNV within *LDLR*. Thus, a more sophisticated and comprehensive molecular screening program that includes an assay for *LDLR* CNVs would be required to accurately diagnose HeFH at the molecular level.

Interestingly, the severity of phenotype found in HeFH patients was significantly associated with the type of mutation. HeFH patients with nonsense mutations or CNVs in *LDLR* had higher untreated LDL cholesterol levels than patients with no detected abnormality or with missense mutations in *LDLR* or *APOB* (Pollex and Hegele, 2007). This example of phenotype-genotype correlation further underlines the potential for large effects due to a CNV within a gene that encodes an important metabolic protein.

Common CNV polymorphism in *LPA* influencing Lp(a) concentration

Plasma Lp(a) concentrations vary over 1000-fold between individuals, but remain stable from birth (Albers and Hazzard, 1974). Heritability estimates for plasma Lp(a) concentrations are ~90%, indicating that this trait is under strong genetic control (Austin et al., 1992; Boerwinkle et al., 1992; Scholz et al., 1999). Studies evaluating Lp(a) and vascular disease have, with some exceptions, shown a positive association (Wild et al., 1997; Craig et al., 1998; Ariyo et al., 2003). Hypothesized mechanisms for Lp(a) and increased vascular disease risk include its involvement in both atherogenic and thrombotic pathways, such as competitive inhibition of plasminogen and interactions with tissue factor pathway inhibitor, platelets, monocytes, smooth muscle cells, endothelial cells, oxidized phospholipids or Lp(a)-associated phospholipase A_2 (Schlaich et al., 1998; Caplice et al., 2001; Discepolo et al., 2006; Kiechl et al., 2007).

Lp(a) is composed of a large polymorphic apo(a) molecule connected via a disulfide bond to the apoB moiety of LDL (Brunner et al., 1993). Apo(a) is a very large protein encoded by the *LPA* gene (MIM 152200), which spans 132 kb on chromosome 6 (6q27). The gene contains a signal peptide region, a catalytic protease domain and five types of 'kringle' domains, which are homologous to the kringle domains found in plasminogen. Apo(a) contains a variable number of kringle IV (KIV) repeats, ranging between 1 and 36 in number, which is the result of genomic duplication or deletion of the two exons and intervening intron that code for KIV (~5 kb at the genomic level) (Lawn et al., 1997). Impaired production and intracellular transport of large protein isoforms encoded by an *LPA* gene with a large number of genomic KIV tandem repeats results in an inverse relationship between the number of genomic KIV repeats and the amount of apo(a) that is assembled and ultimately, the number of Lp(a) particles that can be secreted into plasma. At least 50% of the variation in plasma Lp(a) concentration is due to this KIV copy number polymorphism (Boerwinkle et al., 1992; Rosby and Berg, 2000). A recent study identified strong association between an intronic SNP (rs7770628) and Lp(a) concentrations, suggesting that either linkage between the SNP and the number of KIV repeats exists or that large proportions of Lp(a) plasma concentration variation could be attributed to the two variations (Melzer et al., 2008).

The implication of kringle repeat number for Lp(a) function is not yet understood. To date, identification of Lp(a) KIV repeat genotype requires either digestion of whole genomic DNA followed by pulse-field electrophoresis and Southern analysis (Kraft et al., 1992), or plasma electrophoresis with immunoblotting to identify apo(a) protein isoform size (Kamboh et al., 1991). Both methods are laborious and time consuming; many studies are reported without obtaining the KIV repeat genotype. Furthermore, genome-wide CNV detection methods, such as array-comparative genome hybridization (CGH) or oligonucleotide microarrays, at current resolutions, will have difficulty identifying a CNV of only 5 kb, and are not sensitive enough to differentiate the multiple alleles of the repeat. Other methods to quantify gene dosage, such as quantitative PCR (qPCR), could potentially be used to quickly identify the number of apo(a) KIV repeats in the genomic DNA, although this has not yet been reported, to our knowledge.

Overlap between reported CNVs and genes involved in metabolic syndrome pathways

Recent efforts have been made to measure and map the spectrum of CNVs found in the control population (Redon et al., 2006; Pinto et al., 2007). Collecting CNV data on thousands of individuals was important to define the locations, sizes and frequencies of CNVs. While study participants had no overt disease phenotypes when they provided samples for the study, it is still possible that CNVs in these individuals could modulate susceptibility to development of later-onset complex traits.

Metabolic syndrome (MetS) is a constellation of abnormal phenotypes including dyslipidemia, specifically increased plasma triglycerides and decreased high-density lipoprotein (HDL) cholesterol, dysglycemia, insulin resistance, and abdominal obesity. MetS is extremely common in westernized societies and is associated with 2- to 5-fold increased risk of development of type 2 diabetes and CHD (Pollex and Hegele, 2006). Non-molecular heritability studies suggest a genetic component. Also, mutations have been discovered that cause relatively rare monogenic forms of MetS, and common SNPs have been associated with increased susceptibility to MetS (Pollex and Hegele, 2006).

To calibrate the degree of overlap between reported CNVs and genes with important roles in MetS pathways, a Perl program, *overlap.pl* (available upon request), was written to identify regions of overlap between reported CNVs and genes with potential roles in MetS pathophysiology. Fifty-five genes were selected from a literature review and were separated into the following categories: adipokines, adipose tissue distribution, energy metabolism, lipid metabolism, central regulation and reward sensitivity, and obesity associated. Queries of the NCBI Entrez Gene database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene) were performed to determine the physical location of these genes (NCBI build 36.2). The Toronto Centre of Applied Genomics Database of Genomic Variants (DGV, http://projects.tcag.ca/variation, build hg18v3) (Iafrate et al., 2004) was then searched for all variants, CNVs or indels, that over-

Table 1. Genes implicated in metabolic syndrome containing a CNV or insertion/deletion reported in control populations

Gene name	Gene symbol	Chromo-some	Variation start position (kb)	Size of variation (kb)	Detection method	Region of gene	Gain/loss	Frequency[a]
Adipokines								
Adiponectin	*ADIPOQ*	3	188057	23.9	multiple	3′ 12%	loss	common
			187820	230.2	multiple	3′ 44%	loss	rare
Resistin	*RETN*	19	6661	1126.0	multiple	100%	both	rare
Peroxisome proliferator-activated receptor α	*PPARA*	22	44952	8.4	resequencing	middle 9%	loss	rare
Inflammation								
Interleukin 6	*IL6*	7	22722	165.3	array CGH	100%	both	N/A
Adipose distribution								
Lamin B2	*LMNB2*	19	183	2430.6	multiple	100%	both	common
			2402	0.3	sequencing	middle 1%	loss	rare
Bernardinelli-Seip congenital lipodystrophy 2	*BSCL2*	11	62052	167.3	array CGH	5′ 30%	both	rare
1-acylglycerol-3-phosphate O-acyltransferase 2	*AGPAT2*	9	136968	3304.8	multiple	100%	both	common
Energy metabolism								
Glucocorticoid receptor	*NR3C1*	5	142715	4.4	resequencing	middle 3%	loss	rare
			142618	61.3	array CGH	5′ 27%	loss	rare
			142645	0.1	resequencing	middle 0.1%	loss	rare
Lipid metabolism								
Lipoprotein lipase	*LPL*	8	19784	132.9	array CGH	100%	loss	rare
Hepatic lipase	*LIPC*	15	56630	0.2	resequencing	middle 0.1%	loss	rare
Endothelial lipase	*LIPG*	18	45356	15.9	null SNPs	middle 52%	loss	N/A
			45358	0.3	sequencing	middle 1%	loss	rare
ATP-binding cassette subfamily A member1	*ABCA1*	9	106670	17.9	fosmid end mapping	middle 12%	gain	rare
Glucokinase receptor	*GCKR*	2	27511	189.4	SNP array	100%	gain	rare
CAP-GLY domain containing linker protein 2	*CLIP2*	7	73335	123.0	array CGH	100%	loss	common
Dedicator of cytokinesis 7	*DOCK7*	1	62924	0.3	multiple	middle 0.2%	loss	rare
Central regulation and reward sensitivity								
Serotonin receptor 2C	*HTR2C*	X	112815	2086.5	multiple	100%	both	common
			113903	0.1	resequencing	middle 0.1%	loss	rare
Dopamine receptor D4	*DRD4*	11	201	1112.5	multiple	100%	both	common
Dopamine receptor D5	*DRD5*	4	8573	1351.1	multiple	100%	both	common
Obesity associated								
Leptin	*LEP*	7	127587	131.3	multiple	100%	both	rare
Proopiomelanocortin	*POMC*	2	25141	321.6	array CGH	100%	loss	rare
Fat mass and obesity associated	*FTO*	16	52637	154.7	array CGH	3′ 17%	gain	common
Cytochrome P450 family 11, sub B polypeptide 2	*CYP11B2*	8	143956	36.8	resequencing	5′ 56%	loss	rare

[a] Common: ≥5% in at least one study with a sample size >50; rare: <5% in a study with a sample size >50 or only found in studies with sample sizes <50; N/A: frequency data not available.

lapped the candidate genes. Because the size and number of CNVs detected depends on the technology used, the variant frequency was approximated to either 'common' or 'rare' based upon the frequency the variation was observed in the studies in which it was reported.

Of the 55 genes investigated, 22 (40%) contained a CNV or indel that had been observed in the general population (see Table 1). These CNVs and indels ranged in size from 0.3 kb to 3 Mb. It is not surprising that 40% of genes investigated here overlapped with at least one CNV, since the DGV reports that ~40% (7,242 of 26,091) of MIM genes are overlapped by a CNV (DGV website). However, many of these CNVs may be found in heterozygous carriers for recessive diseases, in individuals at increased susceptibility for development of complex diseases, or in subjects affected with subtle phenotypes. For instance, homozygous or compound heterozygous mutations in 1-acylglycerol-3-phosphate O-acyltransferase 2 (*AGPAT2*, MIM 603100) are causative for an extremely rare metabolic disease called Berardinelli-Seip congenital lipodystrophy (BSCL, MIM 608594). However, three out of 39

unrelated healthy individuals and 16 individuals with previously identified chromosomal imbalances (3/55, ~5%) had complete deletions of *AGPAT2* (Iafrate et al., 2004), suggesting that the frequency of homozygous deletions would be much more common in the population than the prevalence of diagnosed BSCL. Similarly, heterozygous mutations in the gene encoding lipoprotein lipase (*LPL*, MIM 609708), have been reported to have half normal LPL activity, which is associated with elevated plasma triglyceride concentrations (Wilson et al., 1990). One out of 50 European Caucasian controls was identified to carry a complete deletion of the *LPL* gene (de Smith et al., 2007), which is a frequency that greatly exceeds the expectation based on the extraordinarily rare frequency of clinical LPL deficiency. Finally, heterozygous lamin B2 (*LMNB2*; MIM 150341) mutations were found in 5 of 9 patients with acquired partial lipodystrophy (APL, MIM 608709), and not found in a sample of 1,100 control subjects (Hegele et al., 2006), and yet heterozygous deletions and duplications of *LMNB2* have been discovered in up to half of control individuals using array-CGH (de Smith et al., 2007). This again indicates discordance between the high expected frequency of putative disease-causing mutations in *LMNB2* and the very low frequency of clinically ascertained APL.

Why are CNVs within genes that also contain mutations causative for rare phenotypes so common? Potential explanations are biological in nature, such as inaccurate estimates of the rare disease frequency due to subtle or late-onset phenotypes, incompatibility of the homozygous CNV with fetal viability, or additional interacting genetic or environmental factors (Hegele, 2007). Other potential explanations are technical in nature, such as bias within CNV detection methods, or sampling error (Hegele, 2007). Large case-control and family-based association studies are undoubtedly on the horizon to examine the role of these CNVs in modulating susceptibility to metabolic disease (McCarroll and Altshuler, 2007). However, determining the frequency of a CNV remains difficult due to technical issues that will need to be addressed (Scherer et al., 2007), since the power of a study to determine association between a CNV and a phenotype is limited by the sensitivity and specificity of the method used to identify the CNV (McCarroll and Altshuler, 2007).

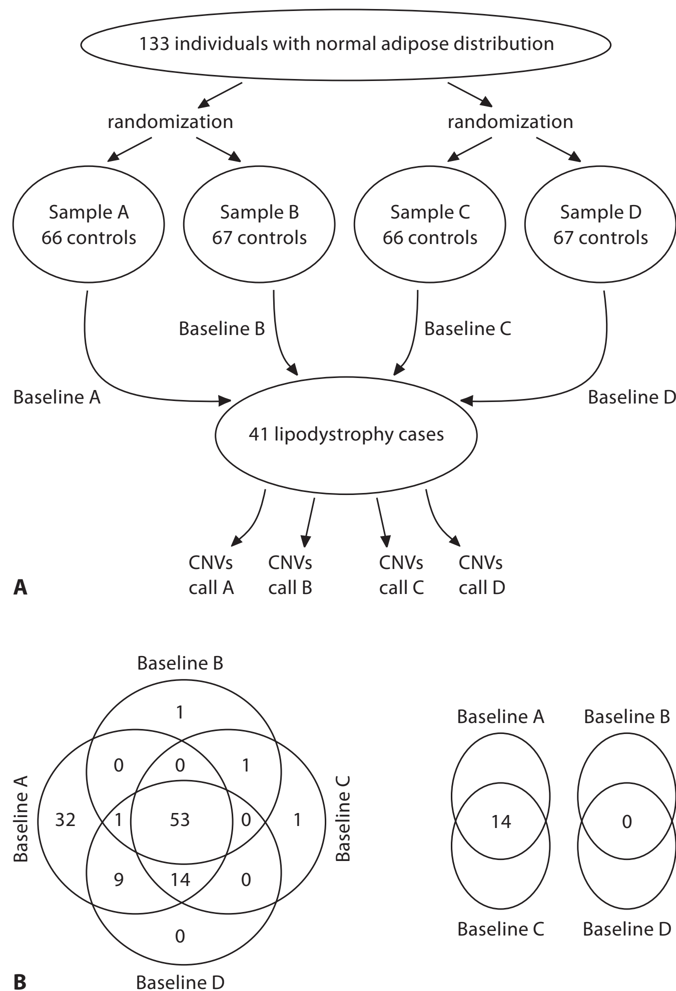

Fig. 1. (**A**) One hundred and thirty-three individuals with no adipose tissue distribution abnormalities were selected as controls for identification of potentially pathogenic CNVs in lipodystrophy patients with no previously detected lipodystrophy-causing mutations. Control subjects were randomized to two groups, and the average SNP intensity for each probe was calculated within each group creating two baselines. The process was then repeated to create a total of four genome-wide sets of baseline intensities. Copy number determination was obtained by comparing an individual's probe intensity to the four baselines using Hidden Markov Models in Partek Genomics Suite v6.3 (St. Louis, MI, USA). (**B**) Venn diagrams representing the overlap of CNVs identified using the four different baselines.

Preliminary screen for rare CNV changes in lipodystrophy patients

Patients with lipodystrophy experience selective or generalized atrophy of adipose tissue. The disruption of lipid metabolism results in a severe form of MetS that also greatly increases the CHD risk (Hegele et al., 2007). A mutation within one of six genes *(LMNA, PPARG, BSCL, AGPAT2, LMNB2* and *CAV1)* is responsible for about half of all lipodystrophy patients (Hegele, clinical observation). Similar to HeFH discussed above, it seems conceivable that a rare, large-effect CNV causing the functional loss of an allele of a gene in which two copies are required to prevent a lipodystrophy phenotype could be identified. However, a preliminary MLPA analysis of *LMNA* (*LMNA* Kit, MRC Holland, Amsterdam, Netherlands) in a sample of 41 lipodystrophy patients with no identified disease-causing mutations (specifically none in *LMNA, PPARG, BSCL, AGPAT2, LMNB2, CAV1* genes) and 40 controls uncovered no *LMNA* CNVs (Lanktree and Hegele, unpublished observations).

As part of a preliminary screening experiment, these same 41 lipodystrophy patients with no known lipodystrophy-causing mutations were genotyped using the Affymetrix GeneChip Human Mapping 500K array set (Santa Clara, CA, USA) (Fig. 1). Because a rare, high-penetrance CNV causative for a dominant monogenic disorder was assumed, we performed a filtering process to remove all CNVs

Table 2. CNVs found in lipodystrophy patients with no previously detected lipodystrophy-causing mutations and not observed in >2 control samples or published controls

Chromosome	~Size (kb)	Start position (kb)	Genes within	Copies	Reports
2	254	133028	*GPR39* (G-protein coupled receptor) *LYPD1* (Ly6/PLAUR domain)	5	2
5	415	154979	None	3	1
6	5300	57436	*GUSBL2* (glucuronidase, beta-like 2) *PRIM2A* (primase, polypeptide 2A) 5 putative loci	3	0
13	125	92534	None	1	1
21	515	35833	*PPP1R2P2* (protein phosphatase) 2 pseudogenes	3	0
21	350	22518	1 pseudogene	1	1
X	all		Triple X syndrome	3	

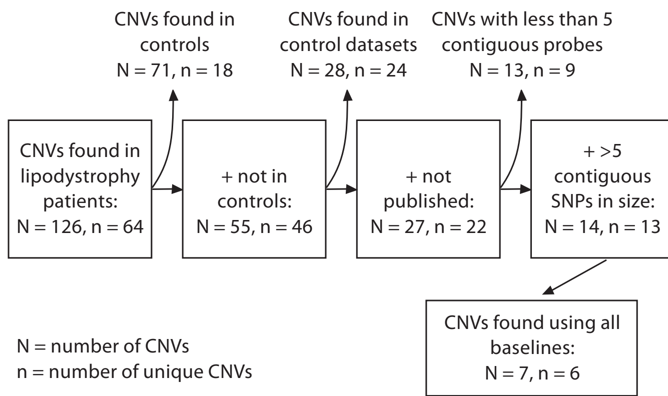

Fig. 2. Filtering algorithm used to enrich for rare, potentially lipodystrophy-causing CNVs. One of the CNVs was found in two lipodystrophy patients who were also sisters.

found in either our control population, a sample of ~2,000 individuals unaffected by lipodystrophy (Marshall et al., 2008), or reported in the DGV (Fig. 2). Six genomic regions ranging in size from 125 kb to 5.3 Mb were identified as being entirely unique for lipodystrophy patients. None of the regions contained genes with known roles in lipodystrophy pathogenesis or metabolic pathways. Each unique CNV region was discovered in only one lipodystrophy proband. Additionally, one lipodystrophy patient was found to be a carrier of three X chromosomes (Table 2).

We do not suspect at this time that any of these regions are necessarily associated with lipodystrophy. The purpose of this preliminary screening experiment was to demonstrate a strategy to identify CNVs that were potentially causative for a rare disorder. If loss of a functional copy of a gene resulted in a lipodystrophy phenotype with high penetrance, then a CNV overlying part of or the entire gene would not be expected to be found in control subjects. Ideally, the discovery of a CNV in lipodystrophy patients but not controls, that also intersects a gene with a known functional metabolic role, or is found in a substantial proportion of lipodystrophy patients, would justify further investigation into a biological, functional role for the CNV in the pathogenesis of lipodystrophy. CNVs not fulfilling these criteria are likely to be simply rare but not necessarily pathogenic CNVs. Study of cosegregation of the CNV with the disease phenotype in extended families would help clarify the issue. One important limitation of this preliminary filtering approach is the potential discounting of a true disease-causing CNV from the candidates, especially if the CNV is not fully penetrant, or requires an interaction with additional genetic or environmental factors, or if the phenotype is present in a mild form and goes undetected in a 'control' subject. Nonetheless, such filtering and evaluation of CNVs might form the basis of a more generic strategy to search for causative CNVs for other rare disorders.

Summary

As methods to identify CNVs increase in their precision and accuracy, the prospect of discovering new genetic causes of both rare Mendelian and common complex diseases will become a reality. However, it will be important to differentiate the methods used to identify a CNV that causes a rare Mendelian metabolic disease in contrast to methods used to identify association between a CNV and susceptibility to a complex metabolic disease. Screening a pool of CNVs discovered in a disease cohort and then removing all CNVs previously identified in control populations will enrich for rare CNVs that are potentially associated with disease, but these CNVs cannot be implicated as causative without additional experimental data. However, a filtering strategy, similar to the one described here, may be appropriate to screen for CNVs that are potentially causative for a rare high-penetrance disease. Alternatively, standard case-control designs, such as those already employed in SNP association studies of complex traits may prove to be ideal for more common CNVs. Large disease cohorts may be re-

quired for study to overcome the reduced power from the low frequency of many CNVs. Additional investigations into CNVs will not only discover CNVs applicable to diagnostic testing, but will uncover new biological mechanisms for human metabolic disease.

Acknowledgements

R.A.H. is a Career Investigator of the Heart and Stroke Foundation of Ontario and holds the Edith Schulich Vinet Canada Research Chair (Tier I) in Human Genetics and the Jacob J. Wolfe Distinguished Medical Research Chair.

References

Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, et al: Medical sequencing at the extremes of human body mass. Am J Hum Genet 80:779–791 (2007).

Albers JJ, Hazzard WR: Immunochemical quantification of human plasma Lp(a) lipoprotein. Lipids 9:15–26 (1974).

Ariyo AA, Thach C, Tracy R: Lp(a) lipoprotein, vascular disease, and mortality in the elderly. N Engl J Med 349:2108–2115 (2003).

Austin MA, Sandholzer C, Selby JV, Newman B, Krauss RM, et al: Lipoprotein(a) in women twins: Heritability and relationship to apolipoprotein(a) phenotypes. Am J Hum Genet 51:829–840 (1992).

Boerwinkle E, Leffert CC, Lin J, Lackner C, Chiesa G, et al: Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. J Clin Invest 90: 52–60 (1992).

Brunner C, Kraft HG, Utermann G, Muller HJ: Cys4057 of apolipoprotein(a) is essential for lipoprotein(a) assembly. Proc Natl Acad Sci USA 90:11643–11647 (1993).

Caplice NM, Panetta C, Peterson TE, Kleppe LS, Mueske CS, et al: Lipoprotein(a) binds and inactivates tissue factor pathway inhibitor: A novel link between lipoproteins and thrombosis. Blood 98:2980–2987 (2001).

Craig WY, Neveux LM, Palomaki GE, Cleveland MM, Haddow JE: Lipoprotein(a) as a risk factor for ischemic heart disease: Metaanalysis of prospective studies. Clin Chem 44:2301–2306 (1998).

Damgaard D, Nissen PH, Jensen LG, Nielsen GG, Stenderup A, et al: Detection of large deletions in the LDL receptor gene with quantitative PCR methods. BMC Med Genet 6:15 (2005).

de Smith AJ, Tsalenko A, Sampas N, Scheffer A, Yamada NA, et al: Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: Implications for association studies of complex diseases. Hum Mol Genet 16:2783–2794 (2007).

Discepolo W, Wun T, Berglund L: Lipoprotein(a) and thrombocytes: Potential mechanisms underlying cardiovascular risk. Pathophysiol Haemost Thromb 35:314–321 (2006).

Hegele RA: Monogenic dyslipidemias: Window on determinants of plasma lipoprotein metabolism. Am J Hum Genet 69:1161–1177 (2001).

Hegele RA: Copy-number variations and human disease. Am J Hum Genet 81:414–415; author reply 415 (2007).

Hegele RA, Cao H, Liu DM, Costain GA, Charlton-Menys V, et al: Sequencing of the reannotated *LMNB2* gene reveals novel mutations in patients with acquired partial lipodystrophy. Am J Hum Genet 79:383–389 (2006).

Hegele RA, Joy TR, Al-Attar SA, Rutt BK: Thematic review series: Adipocyte biology. Lipodystrophies: Windows on adipose biology and metabolism. J Lipid Res 48:1433–1444 (2007).

Holla OL, Teie C, Berge KE, Leren TP: Identification of deletions and duplications in the low density lipoprotein receptor gene by MLPA. Clin Chim Acta 356:164–171 (2005).

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, et al: Detection of large-scale variation in the human genome. Nat Genet 36:949–951 (2004).

Kamboh MI, Ferrell RE, Kottke BA: Expressed hypervariable polymorphism of apolipoprotein (a). Am J Hum Genet 49:1063–1074 (1991).

Kiechl S, Willeit J, Mayr M, Viehweider B, Oberhollenzer M, et al: Oxidized phospholipids, lipoprotein(a), lipoprotein-associated phospholipase a2 activity, and 10-year cardiovascular outcomes: Prospective results from the Bruneck study. Arterioscler Thromb Vasc Biol 27: 1788–1795 (2007).

Kraft HG, Kochl S, Menzel HJ, Sandholzer C, Utermann G: The apolipoprotein (a) gene: A transcribed hypervariable locus controlling plasma lipoprotein (a) concentration. Hum Genet 90: 220–230 (1992).

Lawn RM, Schwartz K, Patthy L: Convergent evolution of apolipoprotein(a) in primates and hedgehog. Proc Natl Acad Sci USA 94:11992–11997 (1997).

Lehrman MA, Schneider WJ, Sudhof TC, Brown MS, Goldstein JL, et al: Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science 227:140–146 (1985).

Lehrman MA, Goldstein JL, Russell DW, Brown MS: Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. Cell 48:827–835 (1987).

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, et al: Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82: 477–488 (2008).

McCarroll SA, Altshuler DM: Copy-number variation and association studies of human disease. Nat Genet 39:S37–42 (2007).

Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, et al: A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet 4:e1000072 (2008).

Pinto D, Marshall C, Feuk L, Scherer SW: Copy-number variation in control population cohorts. Hum Mol Genet 16 Spec No. 2:R168–173 (2007).

Pollex RL, Hegele RA: Genetic determinants of the metabolic syndrome. Nat Clin Pract Cardiovasc Med 3:482–489 (2006).

Pollex RL, Hegele RA: Genomic copy number variation and its potential role in lipoprotein and metabolic phenotypes. Curr Opin Lipidol 18: 174–180 (2007).

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al: Global variation in copy number in the human genome. Nature 444:444–454 (2006).

Reich DE, Lander ES: On the allelic spectrum of human disease. Trends Genet 17:502–510 (2001).

Rosby O, Berg K: *LPA* gene: Interaction between the apolipoprotein(a) size ('kringle IV' repeat) polymorphism and a pentanucleotide repeat polymorphism influences Lp(a) lipoprotein level. J Intern Med 247:139–152 (2000).

Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, et al: Challenges and standards in integrating surveys of structural variation. Nat Genet 39:S7–15 (2007).

Schlaich MP, John S, Langenfeld MR, Lackner KJ, Schmitz G, et al: Does lipoprotein(a) impair endothelial function? J Am Coll Cardiol 31:359–365 (1998).

Scholz M, Kraft HG, Lingenhel A, Delport R, Vorster EH, et al: Genetic control of lipoprotein(a) concentrations is different in Africans and Caucasians. Eur J Hum Genet 7:169–178 (1999).

Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, et al: Human gene mutation database (HGMD): 2003 update. Hum Mutat 21:577–581 (2003).

Tosi I, Toledo-Leiva P, Neuwirth C, Naoumova RP, Soutar AK: Genetic defects causing familial hypercholesterolaemia: Identification of deletions and duplications in the LDL-receptor gene and summary of all mutations found in patients attending the Hammersmith hospital lipid clinic. Atherosclerosis 194:102–111 (2007).

Wang J, Huff E, Janecka L, Hegele RA: Low density lipoprotein receptor *(LDLR)* gene mutations in Canadian subjects with familial hypercholesterolemia, but not of French descent. Hum Mutat 18:359 (2001).

Wang J, Ban MR, Hegele RA: Multiplex ligation-dependent probe amplification of *LDLR* enhances molecular diagnosis of familial hypercholesterolemia. J Lipid Res 46:366–372 (2005).

Wang WY, Barratt BJ, Clayton DG, Todd JA: Genome-wide association studies: Theoretical and practical concerns. Nat Rev Genet 6:109–118 (2005).

Wild SH, Fortmann SP, Marcovina SM: A prospective case-control study of lipoprotein(a) levels and apo(a) size and risk of coronary heart disease in Stanford five-city project participants. Arterioscler Thromb Vasc Biol 17:239–245 (1997).

Wilson DE, Emi M, Iverius PH, Hata A, Wu LL, et al: Phenotypic expression of heterozygous lipoprotein lipase deficiency in the extended pedigree of a proband homozygous for a missense mutation. J Clin Invest 86:735–750 (1990).

Yuan G, Wang J, Hegele RA: Heterozygous familial hypercholesterolemia: An underrecognized cause of early cardiovascular disease. CMAJ 174:1124–1129 (2006).