A gene dosage map of Chromosome 18: A map with clinical utility

Jannine D. Cody, PhD, Erika M. Carter, BS, Courtney Sebold, MS, Patricia L. Heard, BS, and Daniel E. Hale, MD

Purpose: Microarray technology has revolutionized the field of clinical genetics with the ability to detect very small copy number changes. However, challenges remain in linking genotype with phenotype. Our goal is to enable a clinical geneticist to align the molecular karyotype information from an individual patient with the annotated genomic content, so as to provide a clinical prognosis. **Methods:** We have combined data regarding copy number variations, microdeletion syndromes, and classical chromosome abnormalities, with the sparse but growing knowledge about the biological role of specific genes to create a genomic map of Chromosome 18 with clinical utility. **Results:** We have created a draft model of such a map, drawing from our long-standing interest in and data regarding the abnormalities of Chromosome 18. **Conclusion:** We have taken the first step toward creating a genomic map that can be used by the clinician in counseling and directing preventive or symptomatic care of individuals with Chromosome 18 abnormalities. *Genet Med* **2009:11(11):778–782.**

Key Words: *Chromosome 18, chromosome deletions, haploinsufficiency, gene dosage, chromosome abnormality*

During the past 15 years, we have been performing molecular and clinical characterization of individuals with Chromosome 18 abnormalities. The degree of variability seen within our patient population is extensive. In our evaluation of more than 200 individuals with hemizygosity of Chromosome 18q, only two unrelated individuals have identical genomic aberrations.[1,2] In 2002, we reported that 72% of individuals with 18p deletions have breaks at the centromere.[3] However, array comparative genomic hybridization (aCGH) has revealed that even those with so-called "centromeric breaks" have breakpoints that fall into four groups with different genetic content (unpublished data). In addition, individuals with Ring 18 have unique breakpoints on both the p and q arms creating twice the diversity.[4] Given the relatively high number of individuals who have unique regions of hemizygosity, the customary approach of defining genotypically defined "syndromes" is not feasible.

Our ultimate goal is to provide families with a prognostic framework for the care and education of a child with a Chromosome 18 copy number change. Therefore, we will need the ability to piece together an exclusive probabilistic phenotype based on the unique genes involved in a copy number variation (CNV) for each individual. This will require the knowledge of the effect of specific genes, or combinations of genes, in dosage variation.

To that end, we have created a first draft of a gene dosage map of Chromosome 18. This map has been created as a series of custom tracks on the University of California Santa Cruz (UCSC) Genome Browser, which can be aligned with patient's molecular karyotype determined by aCGH. One track includes the gene dosage status of known genes. A second track displays regional information about specific phenotype critical regions within which the exact causative gene has not been identified. These tracks can be focused on a region of an individual person's deletion or duplication. In this way, a clinician could create a prognostic clinical picture of an individual with a Chromosome 18 copy number change. We anticipate that this tool will be a central component for creating an individualized treatment plan for an individual with a Chromosome 18 copy number change.

This way of thinking about chromosome copy number change as the sum of the parts is a more useful approach regardless of whether the copy number change is recurrent or unique. The inclusion of data on phenotype critical regions allows the inclusion of phenotypes that may be the synergistic result of multiple genes in hemizygosity. This approach can accelerate the understanding of and treatments for the clinical consequences of chromosome copy number change because it focuses on the underlying biology. Because this is a first-generation map, the goal of this article is to introduce the idea and format of the gene dosage map rather than present extensive data regarding genotype–phenotype correlations. We think that readers will benefit from learning about our novel approach to the challenge of predicting the phenotype of large chromosome rearrangements based on genotype.

METHODS

Classification of genes

Data from OMIM and medical and scientific literature were reviewed for each of the genes shown on the UCSC Gene track of the UCSC Genome Browser using the March 2006 assembly (www.genome.ucsc.edu).[5] Genes were labeled as haplosufficient if the gene had been identified as:

- Responsible for a recessive condition.
- Hemizygous in multiple CNVs (database of genomic variants).
- Causative of disease by a dominant-negative process.
- Causative of a phenotype that is indistinguishable from wild type in the heterozygous null allele mouse.

Genes were labeled as haploinsufficient if they were associated with disease caused by a single-gene deletion. Regions were labeled as haploinsufficient if they were identified as a

critical region for a phenotype in individuals with 18q- or 18p-.[6] Genes were classified as conditional haploinsufficient if there was evidence that their dosage effects were only manifest in response to an environmental factor (e.g., a drug) or required a second genetic hit to produce the phenotype. Genes were classified as haplolethal if the homozygous null allele mice were not viable. Two regions were identified as such because no individuals had been identified with hemizygosity of the region. Most genes, however, were given an unknown status.

This map is planned to be an evolving data set, so we anticipate that the unknown status genes will eventually be given gene dosage designations. In addition, the status of genes could be changed in light of new information.

Creation of custom tracks

Gene dosage custom tracks were assembled according to the methods described by the UCSC Genome Browser. Custom tracks depicting subject genotype information were created as described previously.[2]

RESULTS

We have created a first draft of a gene dosage map of Chromosome 18. We started this process by reviewing the function of each known RefSeq gene on Chromosome 18 using OMIM and literature searches. More specifically, we determined whether there is any evidence of disease resulting from a hemizygous state. If a listed gene was only hypothetical or if no biological information was available on that gene, it was not included in this initial version of the map. Two hundred fifty-three genes were evaluated for dosage sensitivity and placed into one of the four categories based on their consequences when present in a nondiploid state. These categories, as discussed previously, are as follows:

- Haplolethal (dosage critical—prenatal lethal).
- Haploinsufficient (dosage sensitive).
- Conditional haploinsufficient (conditional dosage sensitive).
- Haplosufficient (dosage insensitive).

Eighty-one genes were determined to be haplosufficient, four were haploinsufficient, and one was a conditional haploinsufficient (Table 1). The effects of the nonhaploid state were unknown for 167 genes.

A custom track was created on the UCSC Genome Browser with this information indicating each of the four categories listed earlier by a different color (Supplemental Digital Content 1, http://links.lww.com/GIM/A84 and http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=SilverSurferTx&hgS_otherUserSessionName=Chr18GeneDosageMap). Haplolethal genes or regions are red, haplosufficient are green, conditional haploinsufficient are yellow, and haploinsufficient are pink. Genes with an unknown role in producing a phenotype when present in an abnormal copy number are shown in gray. A second series of tracks was created using information on critical regions for specific phenotypes that are described in Table 2.[6] Supplemental Digital Content 1, http://links.lww.com/GIM/A84 shows these two sets of tracks for the entire chromosome using these color codes. This track has two regions depicted in red as haplolethal regions. These regions were never found to be hemizygous in the 260 individuals with molecularly characterized abnormalities.[2] Figure 1 shows a zoomed in view of band 18q21.2.

These data can be combined with molecular karyotype data from a patient to determine the clinical ramifications of the chromosome abnormality. Figure 2 shows the molecular karyotype from two subjects aligned with the gene dosage tracks. The top panel shows individual 18q-195C who has a 1.1-Kb interstitial deletion of Chromosomes 18. Although this deletion is very small, it includes a portion of the *TCF4* gene. This gene has been shown to be the cause of Pitt-Hopkins syndrome. Because this gene is known to be disease causing in the hemizygous state, it is classified as being haploinsufficient.[7] Indeed, the phenotype of individual 18q-195C is consistent with Pitt-Hopkins syndrome.

The lower panel of Figure 2 shows the molecular karyotype of subject 18q-71C. This individual has a 5.3-Mb terminal deletion of 18q. The region of hemizygosity is shown aligned with both dosage tracks: the track including genes and the track including regions. The information about the genes and critical regions from our previous work[6] allows us to predict that this individual will have dysmyelination of the brain, a 90% chance of having growth failure with evidence of growth hormone deficiency, a 78% chance of having atretic or stenotic ear canals, and a 25% chance of having a kidney malformation. We are also able to determine that this individual will not be at increased risk for juvenile polyposis or the severe mental retar-

Table 1 Haploinsufficient and conditional haploinsufficient genes on Chromosome 18

Locus	Gene symbol	Gene name	OMIM	Disorder
Haploinsufficient				
18,767,293–18,860,447	*RBBP8*	Retinoblastoma-binding protein 8 isoform a	*604124	Heterozygous knock-out mice have shortened lifespan with multiple tumors, predominantly lymphomas
26,899,940–26,936,386	*DSC2*	Desmocollin-2	125645	Arrhythmogenic right ventricular dysplasia, <30% penetrance
46,810,581–46,865,409	*SMAD4*	Mothers against decapentaplegic, *Drosophila*, homolog of, 4	600993	Pancreatic cancer; polyposis, juvenile intestinal; Juvenile polyposis/hereditary hemorrhagic telangiectasia syndrome
51,040,560–51,406,858	*TCF4*	Transcription factor-4	602272	Pitt-Hopkins syndrome
Conditional haploinsufficient				
3,402,072–3,448,406	*TGIF*, *HPE4*	TG-interacting factor	602630	Holoprosencephaly-4

Table 2 Phenotype critical regions with their molecular definition and penetrance

Phenotype	Critical region borders		Size of region (Mb)	% Penetrance
	Proximal	Distal		
Kidney malformation	70,079,559	73,287,604	3.21	25
Dysmyelination	71,669,548	73,287,604	1.62	100
Growth hormone response failure	71,669,548	73,287,604	1.62	90
Aural atresia	70,983,374	73,287,604	2.30	78

dation and hyperventilation associated with Pitt-Hopkins syndrome because the dosage-sensitive genes on 18q associated with those phenotypes are not within this individual's region of hemizygosity.

DISCUSSION

We have categorized the genes contained on Chromosome 18 into one of the four general categories that will be helpful for predicating clinical significance.

- Haplolethal.
- Haploinsufficient.
- Conditional haploinsufficient.
- Haplosufficient.

These categories represent the spectrum of ways in which genes affect clinical phenotype when present in the haploid state. Those genes that are dosage critical are haplolethal and are required in two (functional) copies. Any variation from that number will be prenatally lethal. Of course proving this will be proving a negative, but as time goes on and as more individuals are genotyped, it will become clear there are a few regions of the genome that are never found in the hemizygous state in a live birth.

At the other end of the spectrum are the genes that are dosage insensitive, i.e., have gene products that are haplosufficient. Dosage-insensitive genes are identified by one of the three mechanisms.

1. These genes can be identified by their presence in regions of normal CNV. The database of genomic variants (http://projects.tcag.ca/variation/) contains a list of more than 6225 structural variations in the human genome larger than 1 Kb that are not associated with an abnormal phenotype.[8]
2. Dosage-insensitive genes can be identified by their involvement in recessive conditions. This is because the definition of a recessive condition is one in which both alleles are nonfunctional to cause disease. By definition, carriers of recessive diseases do not exhibit a phenotype. Because the gene does not cause a phenotype when in the hemizygous state, it is haplosufficient.
3. Dosage-insensitive genes may be identified when a known microdeletion includes a particular gene, however, the phenotype is known to be caused by other gene(s) within that microdeletion region. One example of this is the region of 17p that is duplicated in Charcot-Marie-Tooth disease and deleted in hereditary neuropathy with pressure palsy. This

region contains 21 genes of which only one gene (*PMP22*) is associated with both phenotypes.[9] The other 20 genes in the region do not produce a phenotype when hemizygous or when duplicated and therefore would be classified as haplosufficient.

We hypothesize that most genes on Chromosome 18 are actually haplosufficient. This is supported by the data regarding *PMP22* discussed earlier. In addition, it has been estimated that 90% of mutations are recessive to wild type in *Drosophila*.[10] This leads us to hypothesize that relatively few genes ($\sim$5–10%) will be responsible for the major phenotypic features in a condition like 18q-. There are "officially" only 337 genes on Chromosome 18,[11] so we estimate that only 20 to 30 genes on the entire chromosome will be haploinsufficient. The challenge is to figure out which genes these are.

We hypothesize that there will, in fact, be a spectrum of haploinsufficient phenotypes ranging from those that are "dominant and highly penetrant" to those that are "conditional." We have already identified an example of a dominant and highly penetrant phenotype resulting from haploinsufficiency. We have found that every individual who is hemizygous for a 1.6-Mb region of 18q23 has dysmyelination of the central nervous system.[6] The responsible gene has yet to be identified.

Other haploinsufficient phenotypes will be conditional in nature. By conditional, we mean that there is a second major component, genetic or environmental, required to unmask the haploinsufficient phenotype. For example, an inability to metabolize a certain class of drugs will only be evident if the person is exposed to those drugs. A second type of conditional hemizygosity might be when a second genetic mutation or deletion in another portion of the genome is required before hemizygosity of this particular gene causes a phenotype. An example of this is holoprosencephaly. Hemizygosity of the *TGIF* gene on 18p is not sufficient in and of itself to cause holoprosencephaly. A second genetic "hit" is required.[12]

The two examples illustrated in Figure 2 demonstrate the potential use of this type of genomic annotation. In the top panel, the aCGH data show the individual to be hemizygous for a portion of the *TCF4* gene. Because this gene in the gene dosage track is shown in pink, signifying haploinsufficiency, the clinician knows that it causes a phenotype through haploinsufficiency and that even though this individual has a small deletion it is likely to have clinical consequences. In addition, any of the other genome annotation tracks can also be viewed simultaneously, although this feature is not included in the figure. If for example, the UCSC gene track is also viewed, a clinician can scroll directly to the *TCF4* gene, click on the gene symbol, and be directed to a page with links to a wide variety of sites with information on that gene including OMIM and PubMed. In addition, this gene description page has information about the function of that gene, known diseases associated with that gene, and other summary information on the gene with links to the primary source.

The lower panel in Figure 2 shows the aCGH data from an individual with a much larger region of hemizygosity: a 5.3-Mb deletion in the 18q telomeric region. The gene dosage track shows most of the genes in this region to be gray, meaning that the effects of their hemizygosity are not known. The critical region track, however, indicates that this individual is hemizygous for three phenotype critical regions and the majority of a fourth region. In these critical regions, the single gene responsible for the phenotype has not yet been identified. However, use of these tracks also allows us to take the possible synergistic effects of multiple genes into account. The critical regions
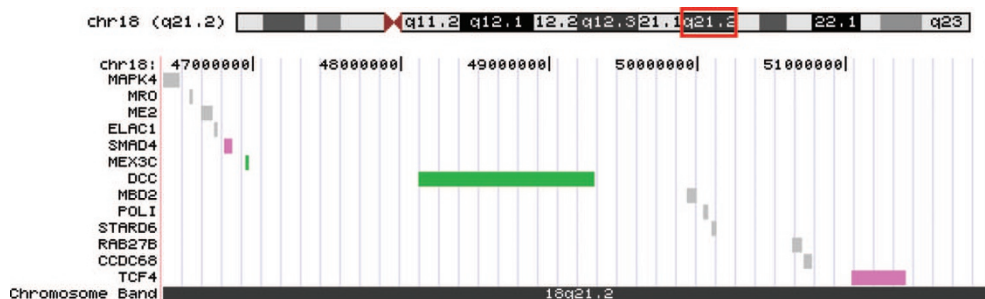
Fig. 1. A zoomed in view from Supplemental Digital Content 1, http://links.lww.com/GIM/A84 of chromosome band 18q21.2 showing the UCSC genes color coded by their dosage sensitivity.
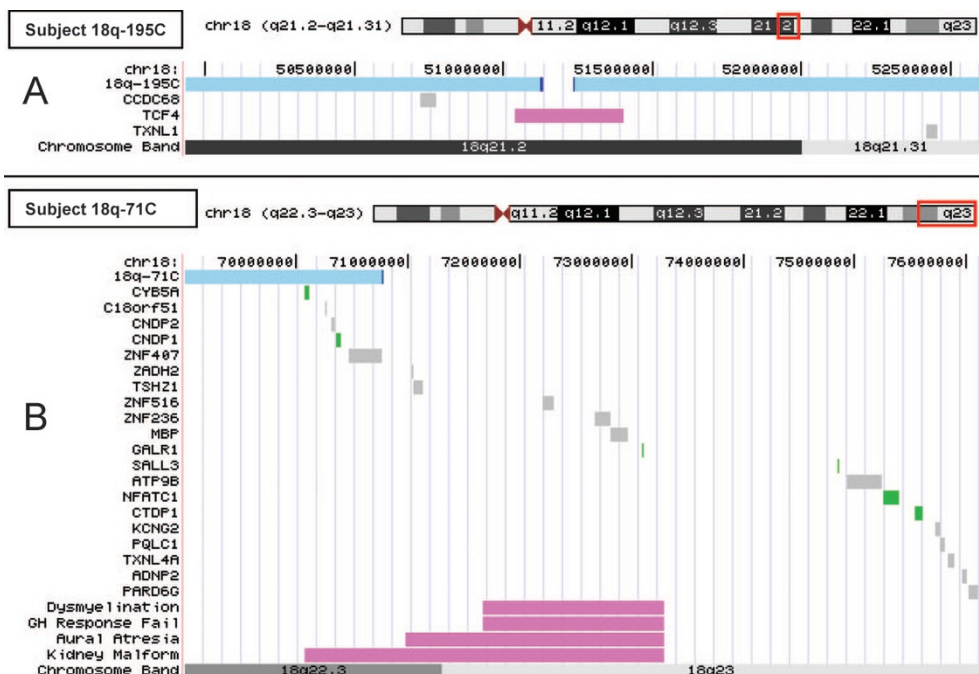

Fig. 2. Two examples of the use of the gene dosage track for the analysis of patient aCGH results. Panel A depicts the data from study participant 18q-195C. The part of their Chromosome 18 that is diploid is shown by the light blue bar, the dark blue segment is their breakpoint region color coded as described for Supplemental Digital Content 1, http://links.lww.com/GIM/A84. Below the patient data are the dosage tracks for the genes in the hemizygous region. Panel B depicts the aCGH data from participant 18q-71C aligned with the genes and the phenotypic critical regions in the hemizygous region; color coded as described in Supplemental Digital Content 1, http://links.lww.com/GIM/A84.

tracks help prevent us from being so reductionist that we overlook the effects of the hemizygosity of multiple genes.

Currently, there is insufficient data about the dosage effects of most genes. However, we identified several critical regions for the phenotypes involved in 18q-. In addition, there are several single-gene disorders (either recessive or dominant haploinsufficient) as well as many CNVs that are known on Chromosome 18. Taken together, this information has allowed us to create a first-generation gene dosage map as a custom track on the UCSC Genome Browser. The information presented in a map format, combined with penetrance estimates,[6] can easily be compared with individual patient's aCGH results to improve prognostic information. As more is learned about the consequences of gene dosage variances, the map will be updated. This map is a first example of using genomic maps in the clinic

to inform appropriate medical care for the one in every 118 babies born every year with a classic chromosome abnormality and the greater number who will be found to have microdeletions by aCGH.[13]

We recognize that this is a completely different way of attacking the problem of understanding conditions with a chromosome copy number basis. Other organizations have been working to create databases and registries of all chromosome abnormalities. We think, however, that this is a self-limiting exercise for several reasons. Although there may be new recurrent microcopy number change conditions yet to be identified, we predict that there are many more unique aberrations than recurring ones. The layering of these overlapping aberrations along with their complex phenotypes across the genomic map will create more confusion than clarity for clinicians looking for

prognostic information. Although this may provide a starting point for a researcher who wants to delve into the role of the individual genes in a region by identifying a potential group of study participants, it makes far more sense to create a consortium of research groups who are each interested in the consequences of CNV within a particular genomic region. However, this requires a completely new way of thinking about the problem of chromosome copy number change. For this reason, we suggest that the strategy of genome annotation will play a larger role than "syndrome hunting" in the eventual prediction of phenotype based on molecular characterization of chromosome abnormalities.

ACKNOWLEDGMENTS

REFERENCES

1. Horbinski C, Carter EM, Heard PL, et al. Molecular and clinical characterization of a recurrent cryptic unbalanced t(4q;18q) resulting in an 18q deletion and 4q duplication. *Am J Med Genet* 2008;22:2898–2904.
2. Heard P, Carter E, Crandall A, Sebold C, Hale DE, Cody JD. High resolution genomic analysis of 18q- using oligo-microarray comparative genomic hybridization (aCGH). *Am J Med Genet A* 2009;149:1431–1437.
3. Schaub RL, Reveles XT, Baillargeon JG, Leach RJ, Cody JD. Molecular characterization of 18p deletions: identification of a breakpoint cluster. *Genet Med* 2002;4:15–19.
4. Stankiewicz P, Brozek I, Hélias-Rodzewicz Z, et al. Clinical and molecular-cytogenetic studies in seven patients with ring chromosome 18. *Am J Med Genet* 2001;101:226–239.
5. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res* 2002;12:996–1006.
6. Cody JD, Heard P, Crandall AC, et al. Narrowing critical regions and determining penetrance for selected 18q- phenotypes. *Am J Med Genet A* 2009;149:1421–1430.
7. Zweier C, Peippo MM, Hoyer J, et al. Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am J Hum Genet* 2007;80:994–1001.
8. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nature* 2004;36:949–951.
9. Inoue K, Dewar K, Katsanis N, et al. The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provided insights into the recent evolution of new genes. *Genome Res* 2001;11:1018–1033.
10. Wilke AOM. The molecular basis of dominance. *J Med Genet* 1994;31:89–98.
11. Nusbaum C, Zody MC, Borowsky ML, et al. DNA sequence and analysis of human chromosome 18. *Nature* 2005;437:551–555.
12. Dubourg C, Bendavid C, Pasquier L, Henry C, Odent S, David V. Holoprosencephaly. *Orphanet J Rare Dis* 2007;2:8.
13. Nielson J, Wohlert M. Chromosome abnormalities found among 34910 newborn children: results from a 13-year incidence study in Arhus, Denmark. *Hum Genet* 1991;87:81–83.