Human Genetics (2018) 137:961–970
https://doi.org/10.1007/s00439-018-1960-6

ORIGINAL INVESTIGATION

Chromosome 18 gene dosage map 2.0

Jannine D. Cody[1,2] · Patricia Heard[1] · David Rupert[1] · Minire Hasi-Zogaj[1] · Annice Hill[1] · Courtney Sebold[1] · Daniel E. Hale[1,3]

Received: 9 August 2018 / Accepted: 14 November 2018 / Published online: 17 November 2018

Abstract
In 2009, we described the first generation of the chromosome 18 gene dosage maps. This tool included the annotation of each gene as well as each phenotype associated region. The goal of these annotated genetic maps is to provide clinicians with a tool to appreciate the potential clinical impact of a chromosome 18 deletion or duplication. These maps are continually updated with the most recent and relevant data regarding chromosome 18. Over the course of the past decade, there have also been advances in our understanding of the molecular mechanisms underpinning genetic disease. Therefore, we have updated the maps to more accurately reflect this knowledge. Our Gene Dosage Map 2.0 has expanded from the gene and phenotype maps to also include a pair of maps specific to hemizygosity and suprazygosity. Moreover, we have revamped our classification from mechanistic definitions (e.g., haplosufficient, haploinsufficient) to clinically oriented classifications (e.g., risk factor, conditional, low penetrance, causal). This creates a map with gradient of classifications that more accurately represents the spectrum between the two poles of pathogenic and benign. While the data included in this manuscript are specific to chromosome 18, they may serve as a clinically relevant model that can be applied to the rest of the genome.

Introduction

The rapidly increasing use of molecular diagnostics is identifying a growing number of people with both small and large genomic copy number changes. However, data regarding the clinical implications of these imbalances lag far behind. Without linking genotype to phenotype, the utility of molecular diagnostics is limited. While some genomic imbalances may have no clinical implications, others may have a serious impact. In between these two extremes lies a wide spectrum of potential outcomes. However, despite this wide range of possible consequences, the genetics community has generally attempted to classify genomic imbalances on a gradient between benign and pathogenic using a scale based on the strength of the evidence for pathogenicity (Richards et al. 2015). The implication is that once all evidence is "very strong" all genomic copy number variations (CNVs) will be either pathogenic or benign. Such a dichotomous classification fails to capture the full breadth of the biology. For example, the data could be very strong that hemizygosity of gene *A* is benign; however, it is causative of disease in the presence of a loss-of-function mutation in gene *B*. Or, gene *A* hemizygosity is causative of disease only with exposure to a particular drug. In both cases, gene *A* hemizygosity alone is benign yet has clinically actionable implications. Conveying the potential consequences is particularly important, because the end goal for each genomic change is the knowledge of the biological consequences of that change and the potential for rectifying those that are adverse.

We are particularly attuned to both genotypic and phenotypic variation due to the non-recurrent nature of most of the chromosome 18 conditions. The vast majority of individuals with chromosome 18 genomic copy number changes have unique regions of genomic variation. This is not a population of people with recurrent CNVs. For example, no two unrelated people with a simple 18q deletion have the same

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00439-018-1960-6) contains supplementary material, which is available to authorized users.

✉ Jannine D. Cody
cody@uthscsa.edu

[1] Department of Pediatrics, The Chromosome 18 Clinical Research Center, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229, USA

[2] The Chromosome 18 Registry and Research Society, San Antonio, TX 78229, USA

[3] Department of Pediatrics, Penn State Milton S. Hershey Medical Center, Hershey, PA 17033, USA

region of hemizygosity, and half of those with 18p deletions have unique deletions (Heard et al. 2009; Hasi-Zogaj et al. 2015). This means that accurately predicting the clinical consequences of an individual's unique deletion or duplication must be based on the genes involved in the imbalance and not merely the chromosome arm within which it lies (e.g., 18q- or 18p-). For this reason, we have created a gene dosage map which annotates each of the known 263 genes on chromosome 18 as well as identifies each region of the chromosome linked to specific phenotypes. The first edition was published in 2009 (Cody et al. 2009a) and is updated annually. The original map included four categories that were mechanistically defined; haplosufficient, haploinsufficient, conditionally haploinsufficient, and haplolethal. However, we now recognize that this classification scheme is as limiting as the dichotomous classifications of benign and pathogenic. In addition, the mechanistic response to a dosage imbalance does not necessarily translate into clinical relevance. For example, mice with a deletion of the *Slc14a1* gene have a urea transport deficiency yet they do not suffer any clinical syndrome, suggesting the existence of compensatory mechanisms (Jiang et al. 2017). While hemizygosity of the *SLC14A1* gene may result in haploinsufficiency in humans at a physiological level the clinical outcome is not significant.

The difficulty in determining where to draw the line between haplosufficient and haploinsufficient; or even between pathogenic and benign became increasingly problematic. Is a gene in hemizygosity that causes an abnormal phenotype benign when this phenotype is 49% penetrant and pathogenic when it is 51% penetrant? Or is it the severity of the phenotype itself that defines the terminology? Is a gene in hemizygosity that causes short stature benign but another one that causes deafness pathogenic? Or is it the age of onset of the abnormal phenotype that is the determining factor? For example is a gene in hemizygosity that causes an adult onset cancer benign but one that causes congenital cataracts pathogenic? These are the questions that prompted us to rethink and re-categorize the terminology and designations from our original gene dosage maps.

We have modified and expanded the classifications of the consequences of abnormal gene dosage to emphasize penetrance and the probability of an abnormal clinically relevant phenotype. This moves away from the binary classification of pathogenic or benign towards a classification system that more adequately reflects the biologic variability. These changes make the second generation genome dosage map more clinically relevant to patients and their healthcare providers.

In addition to describing a more nuanced classification scheme of genomic variants, we will also present several examples of re-evaluation of disease mechanisms in the context of our own data. The updated information on the consequences of abnormal gene dosage has informed the creation of a more nuanced and clinically relevant gene dosage map for chromosome 18 to serve as a model for the rest of the genome.

Materials and methods

The Chromosome 18 Gene Dosage Maps are visualized as custom tracks within the University of California Santa Cruz Genome Browser. The information used to create and update these maps is derived from multiple sources. We utilize the genotype and phenotype data from our cohort of over 650 individuals with chromosome 18 genomic copy number changes. This longitudinal study, now in its 27th year, has been approved by the University of Texas Health Science Center at San Antonio's Institutional Review Board. All study participants are involved in the informed consent process which is appropriately documented. The molecularly defined chromosome 18 copy number changes for each participant were determined by high resolution microarray as previously described (Heard et al. 2009).The clinical consequences and genotype–phenotype correlations have been described in numerous publications and were recently reviewed (Hasi-Zogaj et al. 2015; Sebold et al. 2015; Carter et al. 2015; O'Donnell et al. 2015; Cody et al. 2015). In addition, we use data derived from scientific publications and public databases to inform our classifications. Of particular importance is the Database of Genomic Variation (DGV) (MacDonald et al. 2014) which identifies regions of CNV in control populations thereby eliminating them from high penetrance classifications.

Like the original version, there are two types of data in this second version of the chromosome 18 gene dosage maps: genes and phenotype regions. Each gene on chromosome 18 has been classified into one of the seven classes for hemizygosity and six for suprazygosity. We use the term suprazygosity to combine data on individuals with trisomy 18 as well as tetrasomy 18p. In addition, phenotypes linked to specific regions are classified into five hemizygosity classes and four suprazygosity classes. Phenotype regions are those regions of the chromosome linked to a specific disease or phenotype but for which the causative gene(s) has not yet been identified. These regions range from SNPs to 24 Mb in size.

The classifications for the gene and phenotype regions and the data sources and rationale for each one are shown below. Data from human studies carries more weight than those from animal studies. In particular, data from our own studies provides the most definitive data especially with regard to defining phenotype regions. Recognizing that some conditions have variable expressivity, the presence of any aspect of the associated phenotype is considered to be

evidence of the condition in question. In many cases, however, data from animal studies are the only source of information. Providing the data that are known is more informative than proving no data. The fact that the classifications are displayed using the UCSC Genome Browser means that users can also access other data tracks viewed in parallel, such as ExAC or DECIPHER or OMIM.

Gene hemizygosity classes

1. No clinical effect due to hemizygosity was determined using one or more of these sources.
    a. There is a measurable effect in humans due to hemizygosity but without adverse clinical significance. This could be a blood analyte that is consistently low but still within the normal range.
    b. The DGV shows genomic hemizygosity in more than one individual or this gene was shown to be homozygously deleted in healthy individuals (Sudmant et al. 2015).
    c. The homozygous knockout mouse has no abnormal phenotype.
    d. The heterozygous knockout mouse has no abnormal phenotype.

2. Risk factor for disease from hemizygosity but only in combination with multiple other genetic or environmental factors. These factors are in all cases are not yet identified; but hemizygosity for this gene is found more often in the affected individuals than in controls. Therefore, the existence of hemizygosity in and of itself likely poses a very small increased risk for disease. An example of a gene classified as a risk factor is *LRRC30*. Hemizygosity of this gene was identified more often in people with autism than in people without autism (Pinto et al. 2010). In addition, deletions as well as duplications of this gene have been identified in healthy individuals (MacDonald et al. 2014). The working hypothesis would be that this gene in concert with several other genetic variants or environmental exposures could cause autism thereby making it a risk factor. In all cases, genes in the risk factor classification were associated with conditions known to be polygenic.
3. Conditional cause of disease from hemizygosity but only in the presence of another specific genetic or environmental risk factor. This classification is clinically relevant, because individuals with hemizygosity for any of these genes have a heightened risk, akin to carrier status for a recessive condition, of which their healthcare providers need to be aware.
    a. The other risk factor could be a mutation in or copy number variation in another gene on another chromosome. In addition, the second genetic change could involve a closely linked gene on chromosome 18. For example, hemizygosity of one copy of the TGIF1 gene results in holoprosencephaly, a structural brain malformation, in about 10% of cases, but only when there is a second mutation or deletion of the TWSG1 gene (Rosenfeld et al. 2010).
    b. The other risk factor could be a mutation in the remaining copy of the gene. This would be a revealed mutation for a recessive disease. For this reason genes associated with recessive disease are not classified as benign but rather as conditional.
    c. The secondary factor leading to an abnormal phenotype could be environmental such as a drug exposure. For example, hemizygosity could lead to an altered ability to metabolize a specific class of drugs. The phenotype only becomes apparent upon exposure to a member of that drug class. Loss-of-function mutations in TYMS can cause a reduced ability to metabolize 5-fluorouracil used in chemotherapy. This leads to potentially increased efficacy but also increased toxicity and complication from this chemotherapy treatment (Balboa-Beltrán et al. 2015).

4. Low penetrance of disease occurring as a result of hemizygosity. This is defined as fewer than 50% of people with hemizygosity who exhibit the abnormal phenotype. This determination is based primarily on our own previously reported genotype–phenotype correlation data.
5. Causal of disease if an abnormal phenotype occurs in at least 50% of the people with hemizygosity of this gene. This determination is based on data from any of three sources:
    a. Our own genotype–phenotype correlation data.
    b. Data from the heterozygous knockout mouse.
    c. Data on the single gene human disease literature.

For example, the *TSHZ1* gene was shown to cause aural atresia (Feenstra et al. 2011). In our cohort, 104 individuals with a deletion including this gene were evaluated and 81 (78%) had at least one ear with atresia making this a highly penetrant or causal gene (Cody et al. 2009b).

6. Haplolethal genes are those that are never found in hemizygosity in a human. This determination is based on the failure to identify a human who is hemizygous and is supported by the knockout mouse finding that hemizygosity of this gene leads to prenatal lethality. At this point in time no such genes have been identified on chromosome 18.

7. Unknown annotation classification for the gene, because no data are available regarding the effect of hemizygosity or heterozygous loss-of-function.

Hemizygosity phenotype region classes

1. The mechanism of disease is not directly related to gene dosage.
    a. An abnormal phenotype in which the genetic mechanism is hypothesized to be recessive inheritance means that by definition a heterozygous carrier is unaffected and is the functional equivalent of hemizygosity. However, once the gene is identified and the mechanism confirmed to be recessive this phenotype region would be eliminated and the associated gene would be designated as "Conditional."
    b. Diseases known to be caused by a dominant negative disease mechanism that would not be applicable to hemizygosity.

2. Low penetrance disease associated regions occurring in fewer than 50% of people with hemizygosity based on these data sources:
    a. Based on our study data identifying critical regions by genotype–phenotype correlation mapping in our cohort of over 650 individuals with chromosome 18 copy number changes.
    b. Based on data from the heterozygous knockout mouse data in the literature with the caveat that these data are not always predictive of the human phenotype. Therefore, these data are used with caution.
    c. Based on human disease data in the literature.

3. Causal of disease if hemizygous with a penetrance of at least 50% based on data from the same sources as the Low Penetrance class.
4. Haplolethal based on a critical region never found in hemizygosity in people.
5. Unknown classifications are when no data are available regarding this phenotype with regard to hemizygosity or heterozygous loss-of-function. These data are usually from GWAS studies.

The distinctions between Low Penetrance and Casual are empirical and based study participant data. Whereas, the classifications of Risk Factor and Conditional are based on what is known about the molecular mechanism of the associated disease. As more is leaned about these conditions and the role these genes play in those phenotypes, these categories will become more probability-based. Clearly there is much to learn about the molecular basis of variable penetrance and expressivity that will inform future versions of this classification scheme.

We are also interested in the consequences of chromosome 18 gene duplications. These include individuals with full or partial trisomy 18 as well as individuals with an isochromsome18p resulting in four copies of the genes on 18p. At this point in time data are just beginning to emerge in the literature on the effects of individual gene duplications. Because there is still not a clear delineation between the consequences of 3 copies and 4 copies of any gene on chromosome 18, we have chosen to use the term "suprazygosity". There is no duplolethal class, because the existence of living individuals with trisomy 18 indicates that there are no individual suprazygous lethal genes.

Gene suprazygosity classes

1. No clinical effect due to suprazygosity of this gene or there is a measurable effect but without clinical significance.
    a. Whole gene copy number variations (CNV) are present in more than one unaffected person.
    b. The transgenic mouse has no abnormal phenotype.

2. Risk factor for disease from suprazygosity but only in combination with multiple other genetic or environmental factors thereby posing a very small increased risk for disease.
3. Conditional cause of disease from suprazygosity but only in the presence of another specific genetic or environmental risk factor.
4. Low penetrance disease occurring in fewer than 50% of people with a gene duplication.
5. Causal of disease if suprazygous with a penetrance of at least 50%.
6. Unknown annotation classification for the gene, because no data are available regarding the effect of suprazygosity or gain of function.

Suprazygosity phenotype classes

1. Mechanism of disease not related to gene dosage.
    a. Recessive inheritance.
    b. Dominant negative disease mechanism.

2. Low penetrance disease occurring in fewer than 50% of people with suprazygosity. These regions are usually identified in someone with a small duplication in our study or in the literature.
    a. Based on our study data identifying critical regions by genotype–phenotype correlation mapping.

b. Based on animal model data in the literature.
c. Based on human disease data in the literature such as from a linkage study.

3. Causal of disease if suprazygous with a penetrance of at least 50% based on the same data sources as the Low Penetrance class.
4. Unknown classification when no data are available regarding this phenotype and suprazygosity such as from a GWAS study.

The annotation of each gene and each phenotype region in the Gene Dosage Maps includes the citations for each data source. Users can select the gene or phenotype region of interest and be linked to a details page with an explanation of the classification and the references used to determine the rationale for the classification. The user can then decide if the evidence is sufficient for the classification. The science supporting these classifications is evolving quickly so classifications may change as newer data are published and each are reviewed and updated at least annually.

Results

The Chromosome 18 Gene Dosage Map 2.0 represents a significant advancement from the original version published in 2009[3]. The original version included two sets of custom tracks (genes and phenotype regions) visualized using the UCSC Genome Browser. Because there is now more information available about the effects of suprazygosity, both the gene track and the phenotype region track have been subdivided into two sets of tracks: one based on hemizygosity and one on suprazygosity. Thus, the current version of the Gene Dosage Map has four separate tracks in total.

In addition, the current classifications are now more outcome based rather than mechanistically based. The purpose of these classifications is to convey the risk of an abnormal phenotype resulting from an abnormal gene copy number. Although the mechanism of disease frequently corresponds with penetrance, any one mechanism can be associated with a range of probability of producing an abnormal phenotype. The current classifications are, therefore, organized by the penetrance of an abnormal phenotype rather than molecular mechanism. This improves the clinical utility of the maps.

The number of genes in each classification for the 263 genes on chromosome 18 is shown in Table 1 (Online Resources). Of note, there are significantly more data on the effects of hemizygosity of specific genes than there are on the suprazygosity of specific genes. There are currently no non-coding genomic elements included is this map, because there are no data on their biological role. The number of phenotype regions for in each classification for both Hemizygosity and Suprazygosity are shown below in Table 2 (Online Resources).

These data are displayed as maps in a series of custom tracks on the UCSC Genome Browser with each element in the track linked to our database. This allows the user to select and then view only the region of interest and query the elements (genes and/or phenotypes) in that region. The penetrance-based classification of genes and phenotypes in any region of choice allows the user to compile a more precise clinical picture of the possible effects of a genomic copy number change. The supporting database includes the rationale for the classification as well as the citations of the work supporting that decision.

Two examples are shown in Fig. 1 for two different individuals both with 18q- and both with regions of hemizygosity of similar size. Most dramatically, the individual whose region of hemizygosity is shown in Panel a has an IQ of 120, while the individual in Panel b has the phenotype of Pitt Hopkins syndrome. In addition, the individual in Panel b is not at risk for growth hormone deficiency, hearing loss or dysmyelination of the brain, all associated with 18q-, but now known to be linked to a much more distal region of the chromosome. These examples serve to illustrate how different the outcomes can be and how identifying the specific hemizygous genes in an individual and knowing the role they each play can support clinical care.

The accuracy of the gene dosage map relies on an accurate understanding of the implications of whole gene deletions. For this reason, it was important for us to review previous data and conclusions regarding the mechanisms by which hemizygosity may lead to disease. In performing this analysis of published data, we found discordant data between various sources as well as data that were inconsistent with our findings. *MC4R* is one such example. Mutations in this gene are thought to be an autosomal dominant monogenic cause of obesity (Vaisse et al. 1998). The majority of disease causing mutations reported within the gene are missense mutations leading investigators to conclude that the mechanism of disease was loss-of-function. However, assays of the functional effects of these mutations showed greater than the 50% reduction expected by a single allele loss-of-function (Govaerts et al. 2005). A functional loss greater than 50% suggests a dominant negative effect. It, therefore, is likely that mutations in *MC4R* leading to obesity are the result of a dominant negative effect and not haploinsufficiency. This is supported by our own data in people with hemizygosity of the *MC4R* gene. In 1999, we reported on 27 individuals with a wide variation in the extent of their 18q terminal deletion and we compared the weight of those whose deletion included the *MC4R* gene with those whose deletion did not include the gene. We found no significant weight difference between the two groups (Cody et al. 1999). More recently, we repeated this comparison. Because our

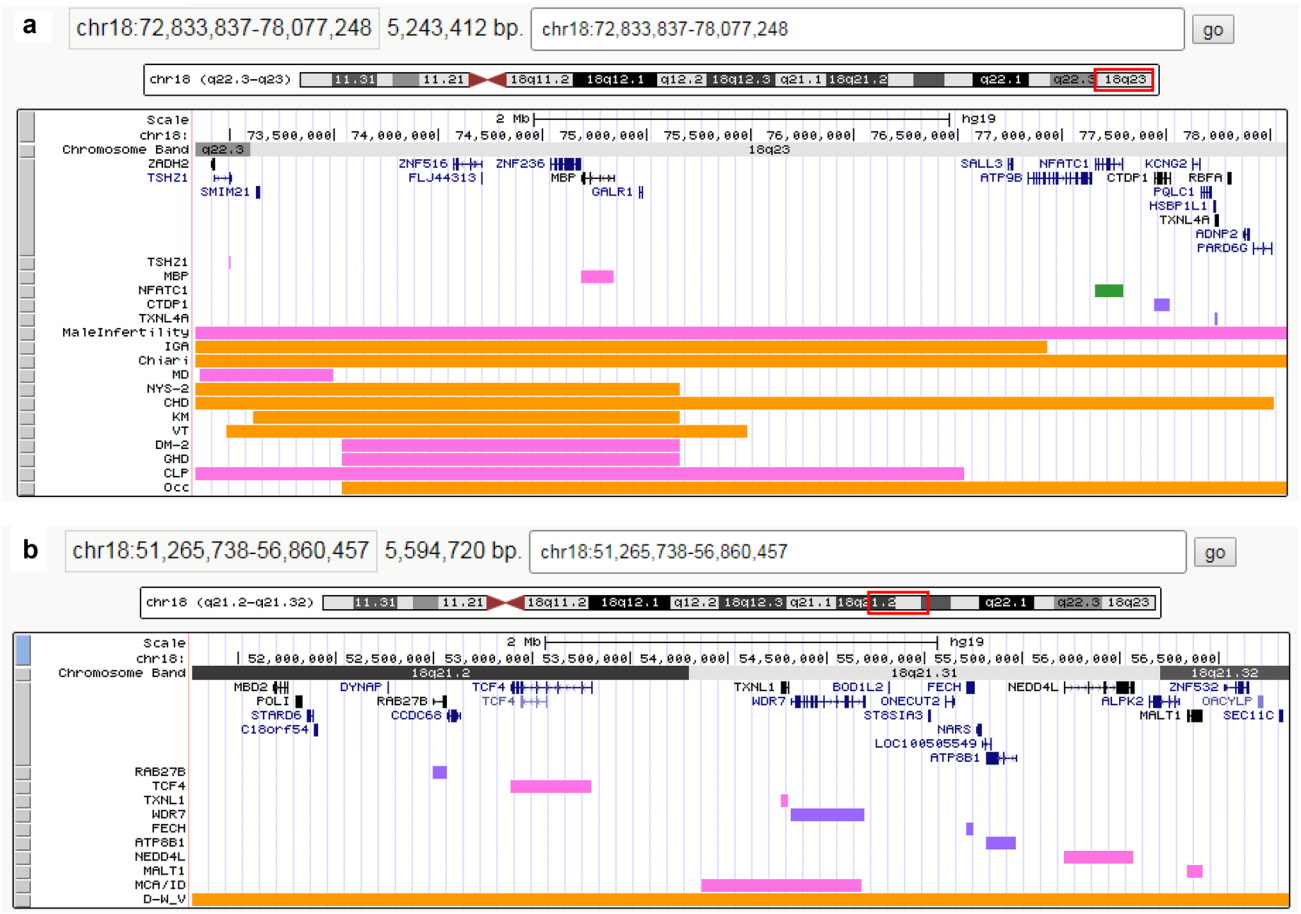

Fig. 1 Gene Dosage Map examples. These figures are partial screen shots from the UCSC Genome Browser that include the Hemizygosity Clinical Dosage Map custom tracks. Panels a and b are for two different individuals with hemizygous regions of similar size. The top of each panel shows the chromosome ideogram with a red box around the portion of the chromosome highlighted in the window below it. The window below includes (from top to bottom) the nucleotide position, the chromosome band, the UCSC Known Genes. The next two tracks are the custom tracks including only the hemizygous genes with clinical implications and the phenotype tracks associated with hemizygosity but for which the causative gene is not known. This particular track combines the genes and the phenotypes and omits the genes with no effect or those with unknown consequences. In **a** the *NFATC1* gene is color coded green, because it is a Risk Factor for scoliosis; the *CTDP1* gene and the *TXNL4A* gene are classified as Conditional and color coded purple as recessively inherited. The *MBP* gene is color coded pink, because it is causal with over 50% penetrance for high frequency neurosensory hearing loss. The Phenotype Regions associated with a penetrance of less than 50% are color coded orange and are IGA deficiency, Chiari malformation, nystagmus, congenital heart disease, kidney malformation, vertical talus and delayed maturation of both occipital lobes. The phenotypes associated with a greater than 50% penetrance are color coded pink and are male infertility, mood disorders, dysmyelination, growth hormone deficiency and a cleft palate abnormality. Each of these items are linked to a details page with additional information and literature citations. **b** Includes no Risk Factor genes (green) and four genes classified as Conditional (purple), *RAB27B, WDR7, FECH* and *ATP8B1* all associated with the recessively inherited conditions. Lastly, four genes are color coded pink, because they are associated with greater than 50% penetrance; *TCF4, TXNL1, NEDD4L* and *MALT1*. There is one phenotype region within this person's hemizygous region which is for a Dandy–Walker malformation

cohort is now much larger and all participants now had high resolution breakpoint determinations we were able to select 14 individuals with terminal deletions of 18q; seven with breakpoints just distal to *MC4R* and seven with breakpoints just proximal to *MC4R*. In this newer analysis the difference between these two groups is the deletion of a single gene: *MC4R*. The BMI percentiles of these two groups were compared (Table 3, Online Resources). These data support the hypothesis that hemizygosity of *MC4R* does not lead to obesity, and therefore, haploinsufficiency is not the mechanism of disease.

We identified two additional examples of genes in which the published disease mechanism predicted a phenotype that we did not see in our large cohort of people with chromosome 18 copy number changes. Therefore, the molecular underpinning of those diseases required review and reconsideration. *ASXL3* is an instructive example. This gene is associated with Bainbridge–Ropers syndrome (BRPS [MIM:

615485]). BRPS is characterized by distinctive facial features, severe developmental disability, and absent speech. Although the mechanism of disease has been postulated to be a dominant loss-of-function (Srivastava et al. 2016) we propose that the mechanism must be a dominant gain of function. The reasoning is as follows; there are 17 molecularly characterized Bainbridge–Ropers patients with truncating mutations in *ASXL3* (Kuechler et al. 2017); none of these same mutations are listed as normal variants in the UniProt database. Yet there are 14 different *ASXL3* truncating mutations in the UniProt database that are classified as normal variants. We would hypothesize that the reason that some truncating mutations lead to disease and others do not is that the disease causing mutations actually lead to a dominant negative effect and the others are merely lead to a loss-of-function. In support of the theory that loss-of-function mutations do not lead to BRPS is the finding that there are 24 individuals in the 1000 Genomes project who have hemizygosity inclusive of the entirety of *ASXL3*. In our own cohort, seven individuals are hemizygous for this gene and none of them have the facial features of Bainbridge–Ropers syndrome and several are verbal. These data lead us to conclude that the hemizygosity of *ASXL3* does not cause an abnormal phenotype and that Bainbridge–Ropers syndrome is likely caused by the dominant negative effects of specific *ASXL3* gene mutations. This conclusion is reflected in the gene's classification in the Gene Dosage Map.

Another example can be found in the *TXNL4A* gene, which is associated with Burn–McKeown syndrome (BMS [MIM: 608572]). BMS is characterized by dysmorphic features, hearing loss, cardiac defects, and choanal stenosis or atresia. This condition is inherited in an autosomal recessive pattern, which implies that the disease causing gene must have homozygous reduction in function mutations (Wieczorek et al. 2014). However, an analysis of genome sequences from over 2500 healthy control individuals found that the 6 most terminal genes on 18q, of which *TXNL4A* is one, were nonessential, because these genes were homozygously deleted in several individuals (Sudmant et al. 2015). Since the *TXNL4A* gene has been identified as dispensable, a mechanism other than homozygous loss-of-function should be considered as the cause of BMS. A review of the genomic variants associated with BMS provides a possible explanation. Whereas one *TXNL4A* allele in BMS may carry any number of loss-of-function variants (nonsense, frameshift, or whole gene deletion), the other allele invariably carries one of the two overlapping promoter deletions. These two "promoter region deletions" are in the promoter region for only some of the transcripts. However, that promotor region is also located within intron 1 of another transcript and in intron 2 of yet another transcript. Thus it is possible that the so-called promoter deletion in individuals with Burn–McKeown actually alters the function or splicing of these other transcripts as well. Because the *TXNL4A* product is a component of the U5 snRNP spliceosome complex which is essential for pre-mRNA splicing, we would hypothesize that either of these two so-called promoter region deletions must result in a gain of function altering the functional activity of the spliceosome. The molecular mechanism of BMS would then be a revealed gain of function mechanism of disease and consistent with the data presented by Wieczorek and coworkers (2014). This hypothesis could reconcile the two sets of data and explain why homozygous whole gene deletions cause no abnormal phenotype (Sudmant et al. 2015) and why BMS has a recessive inheritance pattern. For the purposes of the gene dosage map, this is relevant in providing an accurate risk assessment for individuals with terminal deletions of 18q that include the *TXNL4A* gene. Indeed, there is the likelihood of a patient with 18q- to also exhibit features of Burn–McKeown syndrome if their remaining allele carries a specific genetic variant. This would be an extremely rare event, but this example shows that it is not appropriate to assume that gene associated with an autosomal recessive disease will never have a clinical effect when hemizygous.

To facilitate use of these maps we prepared printed and web-based directions for how to use the gene dosage maps. http://pediatrics.uthscsa.edu/centers/Chromosome18/dosage.asp.

In addition, there are YouTube videos:

How to use the Chromosome 18 Gene Dosage Maps: https://www.youtube.com/watch?v=FrPRccpIBws&t=14s.

Gene Dosage Map 2.0 Examples: https://www.youtube.com/watch?v=ZIqKANvd7E0&t=62s.

Discussion

Medical genetics and in particular cytogenetics has focused on the identification of recurrent genomic syndromes. This important outcome has been achieved by cataloging as many of these genomic copy number changes as possible leading to a compilation of the associated phenotypes. However, the challenge of unique genomic copy number changes that are not recurrent "syndromes" remains. To add to this challenge, some of these genomic copy number changes have clinical ramifications and others do not. To progress toward clinically relevant information for these individuals, we have focused on understanding the contribution, if any that each gene may have play in an abnormal copy number.

Although the ClinGen Dosage Sensitivity Map guidelines (Kearney et al. 2011) suggest a scale for interpretation, that scale defines the strength of the evidence for pathogenic or benign and strives toward a dichotomous classification. The "strength of the evidence" scale is helpful from the perspective of a laboratorian seeking to defend an interpretation. It does, however, not take into account penetrance and

expressivity which are important aspects of clinical interpretation. The scale itself is about the processes and not about the outcome and does not reflect the clinically relevant findings. Only 11% of the genes on chromosome 18 have undergone review by ClinGen compare to 91% on which we have identified sufficient data to assign a classification. Moreover, the database supporting our classifications which is accessible from our Gene Dosage Maps includes only data and citations relevant to gene dosage effects as well as our interpretation of those data. The interpretation makes the maps useful and relevant to non-genetics professionals who can assess the rationale for the classification themselves.

In addition, the ExAC pLI score, which is the probability of a gene being loss-of-function intolerant, has limited utility in this context (Rudefer et al. 2016). Intolerances scores were calculated such that higher positive values indicate greater intolerance (a lower than expected rate of CNVs for that gene) in control populations. Genes intolerant of heterozygous loss-of-function (pLI $\geq$ 0.9) could have an early survival or reproductive disadvantage but it does not mean they could not impact health after reproductive years. This probability could also be impacted by local genetic architecture having nothing to do with the outcome of hemizygosity. This is because genes closer to the centromere and telomere have a higher probability of being in a CNV (Nguyen et al. 2006). Therefore, a high pLI score correlates with a highly penetrant survival limiting gene and can be informative when used in conjunction with other data sources.

Designations such as Risk Factor: or "Conditional" may have limited clinical utility; however, these designations are important for a couple of reasons. First, if for example an individual with a chromosome 18 deletion develops a condition not typically associated with that deletion a clinician can investigate the genes that are classified as Risk Factor or Conditional to gain genetic clues into potential causes. This could help with clarifying a diagnosis as well as the cause in particular in the case of a recessive condition. Secondly, the identification of such patients can further the research into the underlying genetics of potentially polygenic conditions.

Just as the designation of a recessive disease mechanism does not necessarily rule out the possibility of a hemizygosity effect, a dominant disease mechanism does not implicate a gene as having the same clinical effect when present in a hemizygous state. An example of one gene causing two different diseases by dominant mechanisms, one by loss-of-function and another by a gain of function is the *SETBP1* gene. This gene causes Schinzel–Giedion syndrome (SGS [MIM: 269150]) by a dominant negative mechanism (Hoischen et al. 2010) and involves distinct craniofacial features and severe intellectual disability. Known pathogenic mutations fall within four sequences of five consecutive amino acids in exon 4. In contrast, a heterozygous loss-of-function, as occurs in individuals with proximal 18q-, results in different phenotype, namely, significantly delayed expressive language without any craniofacial effects (Cody et al. 2007).

We also found instances of genes being implicated in a disease by hemizygosity that were inconsistent with our own data. For example, *ZNF407* (Zinc finger protein 407) has recently been identified as a having a role in intellectual disability and autism. In a study by Ren et al. (2013) three patients were reported with disruptions of this gene. One had a translocation breakpoint within the gene, while the other two had missense mutations. These authors demonstrated significantly reduced levels of *ZNF407* transcripts from lymphocytes in the affected individuals. These patients had WAIS Intelligence Test scores ranging from 45 to 67 using the Chinese-modified WISC. In addition to intellectual disability, they also had a delay in language development, could not speak clearly, had problems with attention, and did not socialize with their peers. This is a much more severe phenotype than reported in people with terminal deletions of 18q who have deletions of *ZNF407* as well as surrounding genes (Cody et al. 2014). In addition, since the publication of Ren's manuscript, hemizygous deletions of this gene have been identified in healthy individuals (MacDonald et al. 2014). Subsequently, this gene was identified as a recessive cause of intellectual disability (Kambouris et al. 2014) and was identified in a GWAS study as influencing intelligence (Sniekers et al. 2017). Taken together these data suggest that a recessive mechanism is more likely the cause of the phenotype associated with *ZNF407*. The implication is that while hemizygosity alone may not cause these phenotypes there is a small chance that hemizygosity may reveal a mutation in the other allele thereby causing a more severe cognitive deficit than hemizygosity alone.

As the science moves forward and our understanding of the genes on chromosome 18 advances, we will continue to update the Gene Dosage Map. All classifications are based on the currently available data and are subject to revision with the emergence of additional data. Few genes have sufficient data at this time to make a definitive classification. For this reason, Gene Dosage Map users are urged to read the comments regarding any classification to understand how and why the classifications were determined and assume classifications may change as more data are acquired. We are of the opinion that a working hypothesis providing some insight regarding a classification is better than waiting until all definitive data have been collected before a classification is made.

Some types of data are more definitive than others with regard to their implications for a gene dosage classification. For example, the weakest classifications are those made when the only data source are from the Database of Genomic Variation (DGV). Just because non-disease state individuals have been identified with a gene deletion or duplication does not mean the gene never causes disease by a gene dosage

mechanism. It just means it does not always cause disease by that mechanism. For another example, a gene thought to have no clinical effect, because it is hemizygous in a control population could on further evaluation be found to cause a phenotype such as male infertility or another less outwardly obvious or late onset condition. However, given these cautions, the DGV is a very informative tool for determining the likelihood of dosage sensitivity.

One major difference between our approach and that of others is that it is not entirely a reductionist approach to genomic disease management. Because we acknowledge genes whose hemizygosity or suprazygosity are risk factors or have low penetrance effects we are able to facilitate discovery of polygenic phenotypes by identifying one of the components of a polygenic condition the other genetic components of which may be on other chromosomes. In addition, our phenotype map facilities the discovery of polygenic effects caused by closely linked genes. Phenotypes that result from an abnormal gene copy number of one or more genes could be identified as a phenotype region. If no candidate genes are apparent or if there are multiple suggestive candidate genes, it may indicate that the phenotype results only from the additive effect of multiple gene copy number changes. Although OMIM does have a phenotype map, they identify a phenotype region with the original publication and do not narrow or re-define regions based on subsequent information. In addition, they do not include any data from phenotype–genotype correlation studies from people with genomic copy number changes.

In conclusion, these Gene Dosage Maps allow for an individualized clinical interpretation of unique chromosome 18 copy number changes. Equally as important, they demonstrate an approach that could be replicated across the genome, permitting the detection of a CNV to have clinically meaningful, probability-based, implications for health. This approach will move the field closer to a classification scheme that goes beyond the binary choices of pathogenic and benign and closer to clinical guidance and a biological understanding of genomic disease.

Acknowledgements The authors wish to thank the many families who have actively participated in this evolving longitudinal study; many for over 20 years. Support for this work came from the Chromosome 18 Registry and Research Society.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Balboa-Beltrán E, Duran G, Lamas MJ, Carracedo A, Barros F (2015) Long survival and severe toxicity under 5-fluorouracil-based therapy in a patient with colorectal cancer who harbors a germline codon-stop mutation in *TYMS*. Mayo Clin Proc 90(9):1298–1303

Carter E, Heard P, Hasi M, Soileau B, Sebold C, Hale DE, Cody JD (2015) Ring 18 molecular assessment and clinical consequences. Am J Med Genet A 167A(1):54–63

Cody JD, Reveles XT, Hale DE, Lehman D, Coon H, Leach RJ (1999) Haplosufficiency of the melancortin-4 receptor gene in individuals with deletions of 18q. Hum Genet 105(5):424–427

Cody JD, Sebold C, Malik A, Heard P, Carter E, Crandall A et al (2007) Recurrent interstitial deletionsof proximal 18q: a new syndrome involving expressive speech delay. Am J Med Genet A 143A:1181–1190

Cody JD, Carter EM, Sebold C, Heard PL, Hale DE (2009a) A gene dosage map of Chromosome 18: a map with clinical utility. Genet Med 11(11):778–782

Cody JD, Heard PL, Crandall AC, Carter EM, Li J, Hardies LJ et al (2009b) Narrowing critical regions and determining penetrance for selected 18q– phenotypes. Am J Med Genet 149A:1421–1430

Cody JD, Hasi M, Soileau B, Heard P, Carter E, Sebold C, O'Donnell L, Perry B, Stratton RF, Hale DE (2014) Establishing a reference group for distal 18q-: clinical description and molecular basis. Hum Genet 133(2):199–209

Cody JD, Sebold C, Heard P, Carter E, Soileau B, Hasi-Zogaj M et al (2015) Consequences of chromsome18q deletions. Am J Med Genet C Semin Med Genet 169(3):265–280

Feenstra I, Vissers LELM, Pennings RJE, Nillessen W, Pfundt R, Kunst HP et al (2011) Disruption of teashirt zinc finger homeobox 1 is associated with congenital aural atresia in humans. Am J Hum Genet 89:813–819

Govaerts C, Srinivasan S, Shapiro A, Zhang S, Picard F, Clement K (2005) Obesity-associated mutations in the melanocortin 4 receptor provide novel insights into its function. Peptides 26(10):1909–1919

Hasi-Zogaj M, Sebold C, Heard P, Carter E, Soileau B, Hill A, Rupert D, Perry B, Atkinson S, O'Donnell L, Gelfond J, Lancaster J, Fox PT, Hale DE, Cody JD (2015) A review of 18p deletions. Am J Me Genet C Semin Med Genet 169(3):251–264

Heard PL, Carter E, Crandall AC, Sebold C, Hale DE, Cody JD (2009) High resolution genomic analysis of 18q– using oligo-microarray comparative genomic hybridization (aCGH). Am J Med Genet 149A:1431–1437

Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M et al (2010) De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat Genet 42:483–485

Jiang T, Li Y, Layton AT, Wang W, Sun Y, Li M, Zhou H, Yang B (2017) Generation and phenotypic analysis of mice lacking all urea transporters. Kidney Int 91(2):338–351

Kambouris M, Maroun RC, Ben-Omran T, Al-Sarraj Y, Errafii K, Ali R, Boulos H, Curmi PA, El-Shanti H (2014) Mutations in zinc finger 407 [*ZNF407*] cause a unique autosomal recessive cognitive impairment syndrome. Orphanet J Rare Dis 9:80

Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South S, Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee (2011) American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med 13(7):680–685

Kuechler A, Czeschik JC, Graf E, Grasshoff U, Hüffmeier U, Busa T et al (2017) Bainbridge-Ropers syndrome caused by loss-of-function variants in *ASXL3*: a recognizable condition. Eur J Hum Genet 25(2):183–191

MacDonald JR, Zima R, Yuen RK, Feuk L, Scherer SW (2014) The database of genomic variants: a curated collection of structural variation in the human genome. Nucleic Acids Res 42(Database issue):D986–D992

Nguyen DQ, Webber C, Ponting CP (2006) Bias of selection on human copy-number variants. PLoS Genet 2:e20. https://doi.org/10.1371/journal.pgen.0020020

O'Donnell L, Soileau BT, Sebold C, Gelfond J, Hale DE, Cody JD (2015) Tetrasomy 18p: report of cognitive and behavioral characteristics. Am J Med Genet A 167(7):1474–1482

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R et al (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature 15(7304):466 368 – 72.

Ren CM, Liang Y, Wei F, Zhang YN, Zhong SQ, Gu H, Dong XS, Huang YY, Ke H, Son XM, Tang D, Chen Z (2013) Balanced translocation t(3;18)(p13;q22.3) and points mutation in the *ZNF407* gene detected in patients with both moderate non-syndromic intellectual disability and autism. Biochim Biophys Acta 1832(3):431–438

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee (2015 May) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17(5):405–424. https://doi.org/10.1038/gim.2015.30

Rosenfeld JA, Ballif BC, Martin DM, Aylsworth AS, Bejjani BA, Torchia BS et al (2010) Clinical characterization if individuals with deletions of genes in holoprosencephaly pathways by aCGH refines the phenotypic spectrum of HPE. Hum Genet 127:421–440

Ruderfer DM, Hamamsy T, Lek M, Karczewski KJ, Kavanagh D, Samocha KE, Exome Aggregation Consortium, Daly MJ, MacArthur DG, Fromer M, Purcell SM (2016) Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. Nat Genet 48(10):1107–1111

Sebold C, Soileau B, Heard P, Carter E, O'Donnell L, Hale DE, Cody JD (2015) Whole arm deletions of 18p: medical and developmental effects. Am J Med Genet A 167A(2):313–323

Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E et al (2017) Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nat Genet 49(7):1107–1112

Srivastava A, Ritesh KC, Tsan YC, Liao R, Su F, Cao X, Hannibal MC, Keegan CE, Chinnaiyan AM, Martin DM, Bielas SL (2016) De novo dominant *ASXL3* mutations alter H2A deubiquitination and transcription in Bainbridge-Ropers syndrome. Hum Mol Genet 25(3):597–608

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J et al (2015) An integrated map of structural variation in 2,504 human genomes. Nature 526(7571):75–81

Vaisse C, Clement K, Guy-Grand B, Froguel P (1998) A frameshift mutation in human *MC4R* is associated with a dominant form of obesity. Nat Genet 20(2):113–114

Wieczorek D, Newman WG, Wieland T, Berulava T, Kaffe M, Falkenstein D et al (2014) Compound heterozygosity of low-frequency promoter deletions and rare loss-of-function mutations in *TXNL4A* causes Burn-McKeown syndrome. Am J Hum Genet 95(6):698–707