Duplication hotspots, rare genomic disorders, and common disease
Heather C Mefford¹ and Evan E Eichler²,³

The human genome is enriched in interspersed segmental duplications that sensitize approximately 10% of our genome to recurrent microdeletions and microduplications as a result of unequal crossing over. We review the recent discovery of recurrent rearrangements within these genomic hotspots and their association with both syndromic and nonsyndromic diseases. Studies of common complex genetic disease show that a subset of these recurrent events plays an important role in autism, schizophrenia, and epilepsy. The genomic hotspot model may provide a powerful approach for understanding the role of rare variants in common disease.

Addresses
¹ Department of Pediatrics, University of Washington, Seattle, WA 98195, United States
² Department of Genome Sciences, University of Washington, Seattle, WA 98195, United States
³ Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, United States

Corresponding author: Eichler, Evan E (eee@gs.washington.edu)

Introduction
The development of cytogenetic techniques, including high resolution karyotyping and fluorescence in situ hybridization (FISH), in the early 1980s resulted in the identification of microdeletions responsible for Prader–Willi (15q11–q13 deletions) [1] and Smith–Magenis (17p11.2 deletions) [2] syndromes. The term genomic disorder was originally introduced to describe conditions resulting from nonallelic homologous recombination (NAHR) or unequal crossing over between segmental duplications (a.k.a. low copy repeats) [3∗∗]. Over the next decade, continued efforts to fine-map recurrent deletions implicated NAHR for recurrent rearrangements in Charcot–Marie–Tooth disease [4∗], hereditary neuropathy with liability to pressure palsies [5∗], and Prader–Willi [6], Angelman [7], Smith–Magenis [8∗], velocardiofacial [9], Williams–Beuren [10], and Sotos [11] syndromes as well as spinal muscular atrophy [12] and juvenile nephrophthisis type I [13] (Figure 1) to name a few. Molecular diagnosis became possible but relied on first, suspecting a specific disorder based on clinical features and second, using a targeted FISH assay for the chromosomal region to confirm the suspected diagnosis — a ‘phenotype first’ approach.

Advances in technology — most notably the introduction of array comparative genomic hybridization (CGH) and single nucleotide polymorphism (SNP) microarrays — now allow rapid evaluation of many targeted loci or the entire genome for submicroscopic deletions and duplications. A significant advantage of these approaches is that a suspected diagnosis is not necessary before performing the diagnostic test. The application of both targeted and whole-genome technologies to large series of patients with mental retardation (MR) or developmental delay [14∗,15–17,18∗,19], autism [20–22,23∗,24,25], con genital anomalies [26,27∗,28,29], and schizophrenia [30,31,32∗] has had several important consequences. First, the rate of discovery of novel disorders has increased dramatically. Since 2005, 18 new genomic disorders involving 12 regions of the genome have been described, more than doubling the number of disorders described in the previous 20 years (Table 1). Perhaps more importantly, whole-genome approaches have led to a remarkable shift from a ‘phenotype first’ to a ‘genotype first’ definition of genomic disorders. Whereas previously, disorders were described using clinical features, new disorders are described by their genomic rearrangement and clinical features are compared among patients after a common rearrangement is identified. As the diversity of phenotypes evaluated for pathogenic copy number changes expands, so does the phenotypic diversity associated with at least a subset of recurrent rearrangements — in fact, for some of the rearrangements described below, the ‘phenotype first’ approach would have been nearly impossible.

The underlying genomic architecture in each of the genomic disorders identified to date is similar: a stretch of unique sequence (50 kb–10 Mb) flanked by large (>10 kb), highly homologous (>95%) segmental duplications that provide the substrate for NAHR. In 2002, we used these criteria to identify rearrangement ‘hotspots’ — regions predicted to be susceptible to recurrent rearrangement based on the flanking genomic architecture [33∗∗] — and developed a targeted array CGH assay to evaluate copy number variation in both affected and unaffected individuals. An updated map of predicted hotspots and associated disorders is shown in Figure 1; there are now 21 discrete regions of the genome that
Rearrangement hotspots and their associated disease phenotypes. Each chromosome is depicted as a horizontal line with intrachromosomal segmental duplications connected by blue lines. Gold bars represent 'rearrangement hotspots' defined as unique regions (50 kb – 10 Mb) flanked by intrachromosomal segmental duplications >10 kb with >95% sequence identity. Disease-associated deletions and duplications are represented by red and green bars, respectively. Numbers represent rearrangements within hotspot regions as defined above: (1) TAR syndrome, (2) 1q21.1 deletion and (3) reciprocal duplication, (4) juvenile nephronophthisis, (5) 3q29 deletion and (6) reciprocal duplication, (7) spinal muscular atrophy, (8) Sotos syndrome, (9) Williams syndrome and (10) reciprocal duplication, (11) 10q22-q23 microdeletion, (12) Prader–Willi syndrome, (13) Angelman syndrome, (14) duplication 15q11, (15) 15q13.3 microdeletion and (16) reciprocal duplication, (17) 15q24 microdeletion, (18) 16p13.11 deletion and (19) reciprocal duplication, (20) 16p11.2-p12.2 deletion and (21) 16p11.2 deletion and (22) reciprocal duplication, (23) HNPP, (24) CMT1A, (25) Smith–Magenis syndrome, (26) Potter–Lupski syndrome, (27) renal cysts and diabetes (RCAD) syndrome, (28) 17q21.31 microdeletion syndrome, (29) velocardiofacial/DiGeorge/deletion 22q11 syndrome and (30) reciprocal duplication, (31) distal 22q11.2 deletion and (32) reciprocal duplication, and (33) azoospermia. Letters represent NAHR-mediated deletions that occur outside of hotspots (as defined above): (A) Gaucher disease, (B) fascioscapulohumeral dystrophy, (C) congenital adrenal hyperplasia, (D) glucocorticoid-remediable aldosteronism, (E) neurofibromatosis type I microdeletion syndrome, (F) pituitary dwarfism, (G) X-linked ichthyosis, (H) Hunter syndrome, (I) red-green color-blindness, (J) Emery–Dreifuss muscular dystrophy, (K) incontinentia pigmeni, and (L) hemophilia A.
### Table 1

<table>
<thead>
<tr>
<th>Locus</th>
<th>Del or dup</th>
<th>Coordinates (Build 36) and size of critical region</th>
<th>Associated phenotypes</th>
<th>Possible candidate genes</th>
<th>Size and % sequence identity of flanking repeats</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q21.1</td>
<td>del</td>
<td>Chr1: 144.10–144.60 Mb, 500 kb</td>
<td>TAR syndrome: hypomegakaryocytic thrombocytopenia, upper extremity abnormalities ranging from bilateral absent radii to phocomelia; normal intellect</td>
<td>PIAS3, Lix1L</td>
<td>19 kb, &gt;95.0%</td>
<td>[27*]</td>
</tr>
<tr>
<td>1q21.1</td>
<td>del</td>
<td>Chr1: 145.0–146.35, 1.35 Mb</td>
<td>Deletion: variable phenotypes; two groups report mild to severe MR, microcephaly, occasional congenital heart disease; two studies find enrichment of the deletion in schizophrenia</td>
<td>GJA5, GJA8, HYDIN2</td>
<td>281 kb, &gt;99.9%</td>
<td>[30, 31, 51, 52*]</td>
</tr>
<tr>
<td></td>
<td>dup</td>
<td></td>
<td>Duplication: macrocephaly, mild to moderate delays, autistic features; unlike the deletion, has not been seen in schizophrenia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3q29</td>
<td>del</td>
<td>Chr3: 197.4–198.9, 1.5 Mb</td>
<td>Deletion: mild to moderate MR, microcephaly, mild dysmorphic features</td>
<td>PAK2, DLG3</td>
<td>21 kb, &gt;97.1%</td>
<td>[46–48]</td>
</tr>
<tr>
<td></td>
<td>dup</td>
<td></td>
<td>Duplication: mild to moderate MR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10q22–q23</td>
<td>del</td>
<td>Chr10: 81.12–89.07 Mb, 7.95 Mb</td>
<td>Two families reported: deletion carriers have cognitive and behavioral abnormalities of varying severity including: learning disabilities, speech, and language delay, mild developmental delays</td>
<td>NRG3, GRID1, BMPR1, ASNCG, GLUD1</td>
<td>107 kb, &gt;98%</td>
<td>[41]</td>
</tr>
<tr>
<td>15q13.3</td>
<td>del</td>
<td>Chr15: 28.7–30.2 Mb, 1.5 Mb</td>
<td>Deletion: variable phenotypes — mild to severe MR, mild dysmorphism, digital abnormalities, autism; schizophrenia; IGE</td>
<td>CHRNA7</td>
<td>218 kb, &gt;99.4%</td>
<td>[18*, 30, 31, 54–57]</td>
</tr>
<tr>
<td></td>
<td>dup</td>
<td></td>
<td>Duplication: few patients reported, mild to moderate delays; unlike deletion of the same region, has not been reported in schizophrenia or IGE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15q24</td>
<td>del</td>
<td>Chr15: 72.2–73.8 Mb, 1.8 Mb</td>
<td>Mild to moderate MR, high anterior hairline, downslanting PF, long philtrum, digital abnormalities, genital abnormalities, loose connective tissue</td>
<td>MAN2C1, CYP11A1, STRA6</td>
<td>51 kb, &gt;94.0%</td>
<td>[38, 39]</td>
</tr>
<tr>
<td>16p13.11</td>
<td>del</td>
<td>Chr16: 15.4–16.4 Mb, 1 Mb</td>
<td>Deletion: MR, autism, brain abnormalities</td>
<td>NDE1, NTAN</td>
<td>138 kb, &gt;99.0%</td>
<td>[49, 50]</td>
</tr>
<tr>
<td></td>
<td>dup</td>
<td></td>
<td>Duplication: autism, MR; decreased penetrance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16p11.2</td>
<td>del</td>
<td>Chr16: 29.50–30.10 Mb, 600 kb</td>
<td>Deletion: detected in 0.5–1% of individuals with autism; also seen in 0.1% of individuals with psychiatric or language disorders, 0.01% of controls</td>
<td>MAPK3, MAZ, DOC2A, SEZ6L2, HIRIP3</td>
<td>146 kb, &gt;99.4%</td>
<td>[21, 25]</td>
</tr>
<tr>
<td></td>
<td>dup</td>
<td></td>
<td>Duplication: autism, psychiatric or language disorders (0.04%); also seen in 0.03% of population controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16p11.2–p12.2</td>
<td>del</td>
<td>Chr16: 22.0–28.0 Mb, 6 Mb</td>
<td>Severe developmental delay; hypotonia; flat facies, downslanting palpebral fissures, posteriorly rotated ears</td>
<td>Many genes</td>
<td>146 kb, &gt;99.4%</td>
<td>[40]</td>
</tr>
</tbody>
</table>
undergo recurrent rearrangement, resulting in 33 diseases, and at least 10 additional diseases are the result of NAHR in regions of the genome that are flanked by duplications but do not meet our strict definition of a hotspot.

**Mental retardation syndromes**

The majority of the genomic disorders identified before 2006 were characterized by developmental delay, learning disability, and/or MR. Interestingly, the genetic basis for MR is still unknown in well over 50% of clinical cases. Therefore, many studies have been aimed at identifying submicroscopic copy number changes in this population [14,15–17,18,19], and it is now estimated that large microdeletions and microduplications underlie >15% of MR. We note that many potential pathogenic copy number changes are nonrecurrent (i.e. private mutations seen only once) and likely occur by a mechanism other than NAHR since segmental duplications have not been found at the junctions. Although significant in the aggregate, the pathogenicity for any one specific event can be difficult to prove. Here, we focus on those genomic disorders mediated by segmental duplications where the pathogenic significance is unambiguous. Sixteen of the 18 new genomic disorders identified since 2005 are associated with MR (Table 1). Several of these appear to be highly penetrant with recognizable syndromic features.

In 2006, three groups simultaneously reported recurrent microdeletions of chromosome 17q21.31 detected by array CGH [15,18,19], kicking off a flurry of discovery of novel genomic disorders. The 17q21.31 microdeletion, with an estimated prevalence of 1 in 16,000, fits the definition of a classic genomic disorder: the microdeletion has breakpoints in flanking segmental duplications, is always de novo in affected individuals and has never been seen in controls, and patients harboring 17q21.31 microdeletions have very similar phenotypes (Table 1) [15,18,19,34]. Notably, within the same genomic region is an inversion of ~900 kb observed in approximately 20% of individuals of European ancestry [35]. Further emphasizing the importance of regional genomic architecture, the inversion has been found in every parent who transmits a de novo deletion to an affected child and appears to be a prerequisite to facilitate microdeletion [34,36]. The reciprocal duplication has also been reported in one patient with severe psychomotor delay and craniofacial dysmorphism [37]; whether individuals with the reciprocal duplication have syndromic features will require the identification of additional patients.

Reports of recurrent microdeletions in individuals with developmental delay or MR continued steadily throughout 2007 and 2008. Microdeletions of 15q24, although rare, are also highly penetrant. To date, five individuals with overlapping deletions [38,39] and one patient with autism and a larger but overlapping deletion [22] have been reported. As with 17q21.31, all deletions appear to be de novo, and affected individuals have similar facial features in addition to developmental delays (Table 1). Another rare but recognizable syndrome involves deletions of chromosome 16p11.2–p12.2. Ballif et al. [40] reported four individuals (from 8789 analyzed) with severe developmental delays and similar facial features; each had a large deletion sharing the same distal breakpoint at 16p12.2, ranging from 7.1 to 8.7 Mb in size. Deletions of a large hotspot region on chromosome 10q22–q23 are also rare but recurrent. Two families with inherited deletions and one individual with an overlapping deletion have been reported; deletion carriers have varying degrees of cognitive and behavioral abnormalities [41].

The long arm of chromosome 22 is rich with segmental duplications, some of which are responsible for recurrent rearrangements seen in velocardiofacial syndrome, reciprocal 22q11 duplications, and cat-eye syndrome
More recently, recurrent deletions distal to the velocardiofacial syndrome region were reported [44]; most affected individuals had developmental and growth delays and were born prematurely (Table 1). Reciprocal duplications have also been reported and tend to result in milder, more variable phenotypes [45]. Because of the number and density of segmental duplications on 22q, there are several possible rearrangements because of NAHR; many appear to be associated with disease, but collecting information on individuals with the same events is critical to determine features associated with each. Two additional regions for which reciprocal deletions and duplications have been recently reported include 3q29 [46–48] and 16p13.11 [49,50]. Deletions of 3q29 are associated with mild to moderate MR, microcephaly, mild dysmorphic features, and possibly autism; duplications may also be associated with MR but with decreased penetrance. Deletions of 16p13.11 are highly (though not fully) penetrant and have been seen in individuals with autism, MR, dysmorphic features, and brain abnormalities. Individuals harboring duplications tended to have MR, autism, and/or behavioral problems, but the duplication is also seen rarely in controls suggesting decreased penetrance and/or variable expressivity (Table 1).

Non-MR genomic disorders

Although neurocognitive and neurobehavioral diseases appear to be enriched for genomic disorders, this may simply be a result of ascertainment bias. Recent investigations of other diseases suggest that recurrent genomic rearrangements also underlie some disorders that do not include cognitive deficits as a primary phenotype. Array CGH studies of individuals with thrombocytopenia-absent radius (TAR) syndrome found that 30/30 affected probands shared a ~500-kb deletion on chromosome 1q21.1 [27*]. The deletion is not sufficient to cause disease, as it is inherited from an unaffected parent in at least half of cases. It is thought that one or more as-yet unidentified genetic modifiers must play a role. A second disorder described in 2007 is associated primarily with pediatric renal abnormalities and renal cysts and diabetes (RCAD) syndrome [28]. We identified a 1.5-Mb microdeletion of 17q12 encompassing the HNF1B gene in a fetus with severe multicystic dysplastic kidneys. This led us to screen individuals with pediatric renal abnormalities or RCAD syndrome, and we found the identical microdeletion in a subset of patients. The microdeletion appears to be highly penetrant and a frequent cause of early cystic renal disease.

Genomic disorders defying syndromic classification

One of the most intriguing developments over the past two years has been the discovery of at least three new recurrent microdeletions that are enriched in multiple neuropsychiatric diseases but elude syndromic classification. Although each microdeletion was first identified in a series of individuals with similar phenotypes, the application of whole-genome copy number variation analysis to a wider range of neurocognitive disorders has revealed unprecedented phenotypic diversity.

16p11.2 rearrangements

An exciting development in the autism field was the discovery by multiple groups of a recurrent microdeletion of 16p11.2 found in 0.5–1% of affected individuals [21,22,23*,25]. This is one of the most common cytogenetic findings, second only to the 15q11.2 microduplication, for a disorder that has been difficult to tackle from a genetics perspective. However, although significantly enriched in patients with autism, both deletions and duplications are also found in individuals with a psychiatric or language disorder (0.1% and 0.04%, respectively) and in the general population (0.01% and 0.03%, respectively) [25], suggesting extensive variability in expressivity. It is now clear that the deletion is not specific for autism as it is enriched in individuals diagnosed with MR, autism as well as schizophrenia.

Rearrangements of chromosome 1q21.1

Rearrangements of a 1.5-Mb region on 1q21.1, just distal to the deletion found in TAR syndrome, have also been associated with a wide range of phenotypes, including MR and developmental delay [14*,18*,51,52*], schizophrenia [30,31], and congenital heart disease [52*,53]. On the basis of our targeted array of hotspot regions, we identified a de novo deletion in a single patient with developmental delay and MR [18*]. Later, we reported a significant enrichment of both deletions and duplications of the same region in a larger series of patients with developmental delay, MR, and/or congenital anomalies [52*], a finding replicated by Brunetti-Pierri and colleagues [51]. In two large studies of patients with schizophrenia, deletions of 1q21.1 were found in 0.26% of affected individuals compared to 0.02% of population controls [30,31]. At least three of the deletion carriers also had mild cognitive impairment and one had epilepsy. Detailed analysis of deletion breakpoints revealed that individuals with very different phenotypes appear to carry the exact same deletion (Figure 2).

Rearrangements of chromosome 15q13.3

Another microdeletion, first described by Sharp et al. in a series of patients with MR, mild dysmorphic features, and seizures [18*,54], may have even greater phenotype diversity than rearrangements of 1q21.1. Three additional studies have confirmed that the deletion is relatively common in individuals with mild to moderate MR and is also found in a subset of individuals with autism [55–57]. In contrast to the first series of patients reported, very few of the patients in these subsequent studies suffered from seizures. As with 1q21.1 deletions, 15q13.3 deletions are also enriched in individuals with schizophrenia: combined, two studies found 15q13.3 deletions in 0.2% of...
affected individuals compared to 0.01% of unaffected individuals [30,31]. Five of nine deletion carriers in one of the studies [31] also had mild cognitive impairment and one had epilepsy. Finally, yet another large study found enrichment of the same deletion in patients with idiopathic generalized epilepsy (IGE), the most common form of epilepsy [58]. In fact, 15q13.3 microdeletions are more common in IGE (1% of affected individuals) than in MR, autism, or schizophrenia. Again, detailed analysis of breakpoints reveals identical deletions despite highly variable phenotypes (Figure 2). The reciprocal duplication has also been reported in individuals with developmental delays and autistic features [55,57] and rarely in controls, but not in schizophrenia or epilepsy.

**Genomic hotspot model of common and rare disease**

A slight majority of the rearrangements that have been shown to be disease-causing are mediated by segmental duplications. This is simply a consequence of the fact that duplicated sequences promote recurrent rearrangements (Figure 3) requiring far fewer patients and controls to be tested in order to prove pathogenicity when compared to large copy number variants (CNVs) not flanked by segmental duplications. The wide range of phenotypes associated with rearrangements of 16p11.2, 1q21.1, and 15q13.3 points to a common disease mechanism for a wide range of neurocognitive deficits, including autism, MR, epilepsy, schizophrenia, and other psychiatric disorders. While individually, each of these lesions may contribute only 0.1–1% of the total genetic basis of a specific disease, the fact that they influence so many diverse diseases (autism, epilepsy, MR, and schizophrenia) means the overall disease burden is significant, warranting intense scrutiny of these genomic intervals. We propose that these large microdeletions and microduplications are primarily responsible for disease, but the actual specificity of disease is determined by other perhaps more common modifiers — genetic, epigenetic, and environmental. Depending on the severity of the outcome, mildly affected individuals may transmit these alleles to the next generation (explaining both the inherited and de novo aspects), but owing to their high penetrance there is a strong purifying selection operating against the persistence of these alleles in the population. As a result, a high frequency of new mutations or evolutionarily young
mutations, as opposed to ancient inherited mutations, is the primary basis for both common and rare diseases associated with neurodevelopmental (and perhaps other) human diseases.

**Future directions and conclusions**

As we forge ahead in this ‘genotype first’ era of rapid CNV discovery, we should anticipate the need to screen large disease cohorts (10,000–50,000 affected individuals) in order to assess the pathogenicity of other rare CNVs, especially those not flanked by segmental duplications. Some of these numbers may be achieved by leveraging CNV datasets from seemingly disparate disease cohorts (i.e., autism, MR, schizophrenia, and epilepsy). Until such large supracollaborations are established, targeting even smaller hotspots systematically for high-throughput CNV detection may provide a cost-effective way of identifying other pathogenic CNVs. In our work, for example, we identified 107 rearrangement hotspot regions in the human genome, 31% of which (33/107) are now associated with a variety of diseases. Advances in oligonucleotide microarray technology now allow a much larger number of smaller hotspot regions to be assessed for disease association. While genotyping first by array CGH is important, the ability to access extensive phenotypic information for individuals carrying an individual lesion is critical. Simply lumping individuals into categories of disease or controls with no ability to go back to patients (a design common to some genome-wide association studies) is inadequate. As demonstrated by the 15q13.3 microdeletion and its role in epilepsy, detailed clinical information from families provides important clues into other related diseases. High-quality phenotype-genotype correlation is a reiterative process requiring the three-way participation of patient (family), clinician, and researcher. This is something that has been recognized by human geneticists for decades but is worth reiterating as we contemplate the role of duplication hotspots and genomic disorders in common disease.

**Acknowledgments**

We thank Ginger Cheng for assistance in preparation of Figure 1. We apologize to our colleagues whom we could not cite due to the limited number of allowed references. Dr. Eichler is an investigator of the Howard Hughes Medical Institute and is supported in part by the NIH grant HD043569. Dr. Mefford is supported in part by NIH grant HD043376.

**References and recommended reading**

Paper of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


This is a key paper defining ‘genomic disorders’ as conditions that result from DNA rearrangements owing to regional genomic architecture.


CMT1A (duplication of 17p11) and HNPP (deletion of 17p11; Ref. [5]) were the first two genomic disorders described to result from reciprocal rearrangements mediated by segmental duplications.


CMT1A (duplication of 17p11; Ref. [4]) and HNPP (deletion of 17p11) were the first two genomic disorders described to result from reciprocal rearrangements mediated by segmental duplications.


This paper provided the first direct evidence for nonallelic homologous recombination as the mechanism for deletions in Smith-Magenis syndrome.


Authors use the duplication architecture to predict hotspot regions in patients with idiopathic mental retardation and developmental delay; five de novo events were identified including the 17q21.31, 17q13.3, 15q24, 1q21.1, and 17q12 deletion/duplication; all five have since been shown to be recurrent and pathogenic.


This paper shows that 10% of sporadic autism may be due to de novo large copy-number variants.


This paper shows that 10% of sporadic autism may be due to de novo large copy-number variants.


This paper shows that 10% of sporadic autism may be due to de novo large copy-number variants.


One of the first genome-wide studies of copy-number variation among patients which showed that 10% of mental retardation was due to large CNVs which were infrequent in a normal control group.


Genetics of disease


53. Additional detailed assessment of phenotypic details of patients within 1q21.1 deletion shows extreme variability in expressivity including developmental delay, congenital heart defects, cataracts, and/or schizophrenia.


