Nature Reviews Genetics | AOP, published online 30 September 2014; doi:10.1038/nrg3840

DISEASE GENETICS

Searching large and small

> This integrated analysis and forward resequencing approach is useful in both the clinic and research

Copy-number variants (CNVs) have been implicated in many neurodevelopmental disorders; although CNVs can be economically identified using microarrays, it has been difficult to identify the disease-causing genes within them owing to their large size and rarity. By contrast, exome sequencing provides high-resolution mutation information up to the level of single-nucleotide variants (SNVs), but the relatively high cost impairs its application to statistically powered cohort sizes in disease studies. As both CNVs and protein-truncating SNVs result in dosage imbalance for critical genes, a team led by Evan Eichler reasoned that they could improve power in discovering disease-associated genes by systematically integrating CNV and SNV data.

Photodisc

Coe *et al.* constructed a CNV morbidity map from ~20,000 controls and ~29,000 cases of intellectual disability, developmental delay and/or autism spectrum disorders, from which they identified 70 significant CNVs. Notably, the large sample size allowed them to take into account the background level of CNVs and thus identify critical regions that map outside recurrent CNVs mediated by segmental duplications.

They then integrated the CNV data with recently published exome sequencing data and used a joint probability statistic to identify 38 candidate genes with nominally significant enrichment for loss-of-function mutations in cases compared with controls.

Targeted resequencing of ~4,700 cases pinpointed 10 additional genes that were enriched for loss-of-function mutations across CNV and SNV data. Among these genes, a focal *de novo* deletion and five truncating mutations were identified in *SETBP1*, and phenotypic comparison of cases showed that loss of function of this gene is likely to underlie IQ and language deficits, as well as craniofacial dysmorphism.

Importantly, targeted resequencing in large cohorts also allowed adjacent genes within a small significant region of overlap to be discriminated. In the 10p15.3 microdeletion region, only loss-of-function mutations in *ZMYND11* but not the neighbouring *DIP2C* gene were found to be associated with mild intellectual disability, as well as speech and motor delays. Interestingly, cases with truncating mutations in *ZMYND11* also showed complex neuropsychiatric features such as aggression and autistic tendencies, and early diagnosis of this clinical subtype could improve prognosis and outcome.

This integrated analysis and forward resequencing approach is useful in both the clinic and research: not only can new neurodevelopmental syndromes be rapidly discovered, but the underlying genes can also be identified with improved confidence.

Esther Lau

ORIGINAL RESEARCH PAPER Coe, B. P. *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nature Genet.* http://dx.doi.org/10.1038/ng.3092 (2014)