Large-scale discovery of novel genetic causes of developmental disorders

The Deciphering Developmental Disorders Study*

Despite three decades of successful, predominantly phenotype-driven discovery of the genetic causes of monogenic disorders1, up to half of children with severe developmental disorders of probable genetic origin remain without a genetic diagnosis. Particularly challenging are those disorders rare enough to have eluded recognition as a discrete clinical entity, those with highly variable clinical manifestations, and those that are difficult to distinguish from other, very similar, disorders. Here we demonstrate the power of using an unbiased genotype-driven approach2 to identify subsets of patients with similar disorders. By studying 1,133 children with severe, undiagnosed developmental disorders, and their parents, using a combination of exome sequencing3–11 and array-based detection of chromosomal rearrangements, we discovered 12 novel genes associated with developmental disorders. These newly implicated genes increase by 10% (from 28% to 31%) the proportion of children that could be diagnosed. Clustering of missense mutations in six of these newly implicated genes suggests that normal development is being perturbed by an activating or dominant-negative mechanism. Our findings demonstrate the value of adopting a comprehensive strategy, both genome-wide and nationwide, to elucidate the underlying causes of rare genetic disorders.

We established a network to recruit 1,133 children (median age 5.5 years, Extended Data Fig. 1a) with diverse, severe undiagnosed developmental disorders, through all 24 regional genetics services of the UK National Health Service and Republic of Ireland. Among the most commonly observed phenotypes (Extended Data Fig. 1b and Supplementary Table 1) were intellectual disability or developmental delay (87% of children), abnormalities revealed by cranial MRI (30%), seizures (24%), and congenital heart defects (11%). These children are predominantly (≈90%) of northwest European ancestry (Extended Data Fig. 1c), with 47 pairs of parents (4.1%) exhibiting kinship equivalent to, or in excess of, second cousins (Extended Data Fig. 1d and Supplementary Information). In most families (849 of 1,101) the child was the only affected family member, but 111 children had one or more parents with a similar developmental disorder, and 124 had a similarly affected sibling (Supplementary Information). Prior clinical genetic testing would have already diagnosed many children with easily recognized syndromes, or large pathogenic deletions and duplications, enriching this research cohort for less distinct syndromes and novel genetic disorders.

We sequenced the exomes of 1,133 children with developmental disorders and their parents, from 1,101 families, representing 1,071 unrelated children and 30 sibships. We also performed exome-focused array comparative genomic hybridization (exome-aCGH) on the children (n = 1,009) and UK controls (n = 1,013), and genome-wide genotyping on the trios (n = 1,006) to identify deletions, duplications, unparental disomy and mosaic large chromosome rearrangements. From our exome sequencing and exome-aCGH data, we detected an average of 19,811 coding or splicing single nucleotide variants (SNVs), 491 coding or splicing insertions and deletions (indels) and 148 copy number variants (CNVs) per child (Supplementary Information). From analyses of the genotyping array data2 we identified six children with unparental disomy and five children with mosaic large chromosomal rearrangements (Supplementary Information). The SNVs, indels and CNVs were analysed jointly in the following analyses, allowing, for example, the identification of compound heterozygous CNVs and SNVs affecting the same gene.

We discovered 1,618 de novo variants (1,417 SNVs, 114 indels and 87 CNVs) in coding and non-coding regions (Supplementary Tables 2 and 3), of which 1,596 (98.6%) were validated using a second, independent assay, and the remainder were validated clinically. This represents an average of 1.12 de novo SNVs and 0.09 de novo indels in coding or splicing regions per child, which is within the range of similar studies2–11. The distribution of de novo SNVs and indels per child closely approximated the Poisson distribution expected for random mutational events (Extended Data Fig. 2).

We classified 28% (n = 317) of children with probable pathogenic variants (Supplementary Table 4 and ref. 13) in 1,129 robustly implicated developmental disorder genes (published before November 2013), or with pathogenic deletions or duplications. Most of these diagnoses involved de novo SNVs, indels or CNVs (Table 1). Females had a significantly higher diagnostic yield of autosomal de novo mutations than males (P = 0.01, Fisher’s exact test). Among the single-gene diagnoses, most genes linked to developmental disorders (95 out of 148) were only observed once, although eight (ARID1B, SATB2, SYNGAP1, ANKRD11, SCN1A, DRYK1A, STXBP1, MED13L) each accounted for 0.5–1% of children in our cohort (Extended Data Fig. 3). For seventeen of these children we identified two different genes with pathogenic variants, resulting in a composite clinical phenotype.

Analyses that assess the enrichment in patients of a particular class of variation, so-called ‘burden analyses’, both highlight classes of variants for detailed analysis and enable estimation of the proportion of a particular class of variant that is likely to be pathogenic. We observed a significant (P = 0.0004) burden of 87 de novo CNVs in the 1,133 children with developmental disorders compared to 12 in 416 controls (Scottish Family Health Study14) despite most children (77%) having previously had clinical microarray testing (Extended Data Fig. 4).

We used gene-specific mutation rates that account for gene length and sequence context15 to assess the burden of different classes of de novo SNVs and indels (Supplementary Information). We observed no significant excess of any functional class of de novo SNVs or indels in

| Table 1 | Breakdown of diagnoses by mode and by sex |
|-----------------|-----------------|-----------------|-----------------|
|                | Female (%)      | Male (%)        | Total (%)       |
| Undiagnosed    | 383 (69.6)      | 433 (74.3)      | 816 (72.0)      |
| Diagnosed      | 167 (30.4)      | 150 (25.7)      | 317 (28.0)      |
| De novo mutation| 124 (22.5)      | 80 (13.7)       | 204 (18.0)      |
| Chr X*         | 24 (4.4)        | 5 (0.9)         | 29 (2.6)        |
| Autosomal*     | 100 (18.2)      | 75 (12.9)       | 176 (15.5)      |
| Autosomal dominant†| 9 (1.6)    | 11 (1.9)        | 20 (1.8)        |
| Autosomal recessive| 20 (3.6)  | 26 (4.5)        | 46 (4.1)        |
| X-linked inherited| 1 (0.2)    | 19 (3.3)        | 20 (1.8)        |
| UPD/mosaicism  | 4 (0.7)         | 6 (1.0)         | 10 (0.9)        |
| Composite      | 9 (1.6)         | 8 (1.4)         | 17 (1.5)        |
| Total          | 550             | 583             | 1,133           |

* Lists of participants and their affiliations appear at the end of the paper.
autosomal-recessed developmental-disorder-linked genes (Extended Data Fig. 5), suggesting that few of these mutations are causally implicated. By contrast, we observed a highly significant excess of all ‘functional’ classes (coding and splice site variants excepting synonymous changes) of de novo SNVs and indels in the dominant and X-linked developmental-disorder-linked genes (Extended Data Fig. 5) within which de novo mutations can be sufficient to cause disease. Not all protein-altering mutations in known dominant and X-linked developmental disorder genes will be pathogenic, and these burden analyses inform estimates of positive predictive values for different classes of mutations. The remaining genes (that is, those not linked to developmental disorder) in the genome also exhibit a more modest, but significant, excess of functional, but not silent, de novo SNVs and indels (Extended Data Fig. 5).

We observed 96 genes with recurrent, functional mutations (Fig. 1a), a highly significant excess compared to the expected number derived from simulations (median = 55; Supplementary Information). This enrichment is even more pronounced (observed, 29; expected, 3) for recurrent loss-of-function mutations (Fig. 1b). Among undiagnosed children, we observed an excess of 22 genes (observed: 45, expected: 23) with recurrent functional mutations (Fig. 1a) and an excess of 8 genes (observed, 9; expected, 1) with recurrent loss-of-function mutations (Fig. 1b), implying that an appreciable fraction of these recurrently mutated genes are novel developmental-disorder-linked genes.

To identify individual genes enriched for damaging de novo mutations (Supplementary Information), we tested for a gene-specific over-abundance of either de novo loss-of-function mutations or clustered functional de novo mutations in 1,130 children (excluding one twin from each of three identical twin pairs). To increase power to detect genes associated with developmental disorder, we also meta-analysed our data with published de novo mutations from 2,347 developmental disorder trios with intellectual disability4,9, epileptic encephalopathy3, our data with published genes associated with developmental disorder, we also meta-analysed from each of three identical twin pairs). To increase power to detect de novo abundance of either mutations (Supplementary Information), we tested for a gene-specific over-enrichment is even more pronounced (observed, 29; expected, 3) for recurrent loss-of-function mutations (Fig. 1a) and an excess of 22 genes (observed: 45, expected: 23) with recurrent functional mutations (Fig. 1b). Among undiagnosed children, we observed an excess of 8 genes (observed, 9; expected, 1) with recurrent loss-of-function mutations (Fig. 1b), implying that an appreciable fraction of these recurrently mutated genes are novel developmental-disorder-linked genes.

To identify individual genes enriched for damaging de novo mutations (Supplementary Information), we tested for a gene-specific over-abundance of either de novo loss-of-function mutations or clustered functional de novo mutations in 1,130 children (excluding one twin from each of three identical twin pairs). To increase power to detect genes associated with developmental disorder, we also meta-analysed our data with published de novo mutations from 2,347 developmental disorder trios with intellectual disability4,9, epileptic encephalopathy3, autism4–8,10, schizophrenia3, or congenital heart defects11 (the ‘meta-DD’ data set). These analyses (Fig. 2) successfully re-discovered 20 known genes linked to developmental disorder at genome-wide significance (P < 1.31 × 10⁻⁶, a Bonferroni P value of 0.05 corrected for 38,504 tests (Supplementary Information)). Thus, despite the broad phenotypic ascertainment in these data sets, we can robustly detect developmental-disorder-linked genes solely on statistical grounds.

To increase our power to detect novel genes linked to developmental disorder, we repeated the gene-specific analysis described above excluding the 317 individuals with a known cause of their developmental disorder. In this analysis the statistical genetic evidence was integrated with phenotypic similarity of patients, available data on model organisms and functional plausibility. We identified twelve novel disease genes with compelling evidence for pathogenicity (Table 2), nine of which exceeded the genome-wide significance threshold of 1.36 × 10⁻⁶ (Supplementary Information), with the remaining three genes (PCGF2, DNM1 and TRIO) just below this significance threshold. The two children with identical Pro65Leu mutations in PCGF2, which encodes a component of a Polycomb transcriptional repressor complex, share a strikingly similar facial appearance representing a novel and distinct dysmorphic syndrome. DNM1 was previously identified as a candidate gene for epileptic encephalopathy3. Two of the three children that we identified with DNM1 mutations also had seizures, and a heterozygous mouse mutant manifests seizures7. In addition to two de novo missense SNVs in TRIO, we identified an intragenic de novo 82-kilobase (kb) deletion of 16 exons. For several of these novel developmental-disorder-linked genes, the meta-DD analysis increased the significance of enrichment. For example, a total of five de novo loss-of-function variants in POGZ were identified, two from our cohort, two from recent autism studies6–8 and one from a recent schizophrenia study7. We also identified six genes with suggestive statistical evidence of being novel genes associated with developmental disorder, defined as having a P value for mutation enrichment less than 1 × 10⁻⁴ and being plausible from a functional perspective (Extended Data Table 1). We anticipate that most of these genes will eventually accrue sufficient evidence to meet the stringent criteria we defined above for declaring a novel developmental-disorder-linked gene.

Notably, we observed identical missense mutations in unrelated, phenotypically similar patients for four of these novel developmental-disorder-linked genes (PCGF2, COLA4A3BP, PPP2R1A and PPP2R5D), and for a fifth gene, BCL11A, we identified highly significant clustering of non-identical missense mutations (Fig. 3). We hypothesize that the mutations in some of these genes may be operating by either dominant-negative or activating mechanisms. This hypothesis is supported by previous functional evidence for several of the mutated amino acids. The three identical Ser132Leu mutations in COLA4A3BP, which encodes an intracellular transporter of ceramide, remove a serine that when phosphorylated downregulates transporter activity from the ER to the
Golgi17, presumably resulting in intra-cellular imbalances in ceramide and its downstream metabolic pathways. The two mutated amino acids (Arg182Trp and Pro179Leu) in PPP2R1A, which encodes the scaffolding A subunit of the protein phosphatase 2 complex, have been previously identified as sites of driver mutations in endometrial and ovarian cancer18. It has previously been shown that mutating either of these two residues results in impaired binding of B subunits of the complex18. Intriguingly, PPP2R5D encodes one of the possible B subunits of the same protein phosphatase 2 complex, suggesting that the clustered missense mutations (Pro201Arg and Glu198Lys) in this gene may similarly perturb interactions between subunits of this complex. Further functional studies will be required to confirm this hypothesis.

We assessed transmission biases of potentially pathogenic inherited SNVs in our probands (Supplementary Information) and observed a genome-wide trend (P = 0.015) towards over-transmission to probands of very rare (minor allele frequency (MAF) <0.0005%) loss-of-function variants, but not damaging missense variants. We also observed a 1.8-fold enrichment (P = 0.04) of rare (MAF <5%) biallelic loss-of-function variants (Supplementary Table 5) among probands without a likely dominant cause of their disorder, compared to those with either a diagnostic de novo mutation or an affected parent. Again we saw no enrichment in biallelic damaging missense variants (Extended Data Table 2), consistent with a similar observation in children with autism19. These observations suggest that although inherited loss-of-function variants (both monoallelic and biallelic) are probably contributing to developmental disorder in our patients, much larger sample sizes will be required to pinpoint specific developmental-disorder-linked genes in this way.

To direct future, detailed functional experiments on the developmental role of a subset of candidate genes from this study we used two approaches. First, knockdown-induced phenotypes were recorded in early zebrafish development. Second, we performed a systematic review of perturbed gene function in human, mouse, Xenopus, zebrafish and Drosophila. In both approaches the animal phenotypes were compared to those seen in individuals in our cohort.

We undertook an antisense-based loss-of-function screen in zebrafish to assess 32 candidate developmental-disorder-linked genes with de novo loss-of-function, de novo missense or biallelic loss-of-function variants from exome sequencing (Supplementary Information and Supplementary Table 6). These candidate genes corresponded to 39 zebrafish orthologues. Knockdowns of these zebrafish genes were repeated at least twice and all morpholinos were co-injected with tp53 morpholino to eliminate off-target toxicity. Successful knockdown of the targeted messenger RNA could be confirmed using polymerase chain reaction with reverse transcription (RT–PCR) for 82.4% of genes (28 out of 34), and 9 out of 11 (82%) of genes that were tested gave an equivalent phenotype when knocked down by a second, independent morpholino. Knockdown of at least one or a pair of zebrafish orthologues of 65.6% of the zebrafish morphants to those of the patients with de novo mutations or biallelic loss-of-function variants in the orthologous genes (Extended Data Table 3). Eleven out of twenty-one (52.4%) of the genes were categorized as strong candidates based on phenotypic similarity (Fig. 4a). Seven out of eleven were potential microcephaly genes, the knockdown of which in zebrafish gives significant reductions in both head measurements and neural tissue (Fig. 4b and Supplementary Information). Six out of twenty-one (28.6%) genes resulted in severe

Table 2 | Novel genes with compelling evidence for a role in developmental disorder

| Evidence       | Gene   | De novo DDD (missense, LOF) | De novo meta (missense, LOF) | P value | Test     | Mutation clustering | Predicted haploinsufficiency (%) |
|----------------|--------|-----------------------------|-----------------------------|---------|---------|---------------------|---------------------------------|
| De novo enrichment | COL4A3BP | 3 (3,0)                    | 5 (5,0)                     | 4.10 × 10⁻¹² | Meta | Yes | 14.7 |
|                 | PPP2R5D | 4 (4,0)                    | 5 (5,0)                     | 6.01 × 10⁻¹² | DDD  | Yes | 19.7 |
|                 | ADNP   | 4 (0,4)                    | 5 (0,5)                     | 4.59 × 10⁻¹¹ | Meta | No  | 9.8  |
|                 | POGZ   | 2 (0,2)                    | 5 (0,5)                     | 4.31 × 10⁻¹⁰ | Meta | No  | 30.0 |
|                 | PPP2R1A | 3 (3,0)                    | 3 (3,0)                     | 2.03 × 10⁻⁸  | DDD  | Yes | 23.5 |
|                 | DOXX   | 4 (2,1)                    | 5 (3,2)                     | 2.26 × 10⁻⁷  | DDD  | No  | 12.7 |
|                 | CHAMP1 | 2 (0,2)                    | 3 (0,3)                     | 4.58 × 10⁻⁷  | Meta | No  | 52.9 |
|                 | BCL11A | 3 (3,0)                    | 4 (3,1)                     | 1.03 × 10⁻⁶  | DDD  | Yes | 0.6  |
|                 | PURA   | 3 (1,2)                    | 3 (1,2)                     | 1.14 × 10⁻⁶  | DDD  | No  | 9.4  |
| De novo enrichment plus additional evidence | DNM1 | 3 (3,0) | 5 (5,0) | 1.43 × 10⁻⁶ | Meta | No | 13.5 |
|                 | TRIO   | 2 (2,0)                    | 7 (7,0)                     | 5.16 × 10⁻⁶  | Meta | Yes | 25.7 |
|                 | PCGF2  | 2 (2,0)                    | 2 (2,0)                     | 1.08 × 10⁻⁵  | DDD  | Yes | 37.7 |

The table summarizes the 12 genes with compelling evidence to be novel developmental-disorder-linked genes. The number of unrelated patients with independent functional or loss-of-function (LOF) mutations in the Deciphering Developmental Disorders (DDD) cohort or the wider meta-analysis (meta) data set including DDD patients is listed. The P value is the minimum P value from the testing of the DDD data set and the meta-analysis data set. The data set that gave this minimal P value is also reported. Mutations are considered to be clustered if the P value of clustering of functional SNVs is less than 0.01. Predicted haploinsufficiency is reported as a percentile of all genes in the genome, with 0% being highly likely to be haploinsufficient and 100% very unlikely to be haploinsufficient, based on the prediction score described in ref. 26 updated to enable predictions for a higher fraction of genes in the genome. During submission, a paper was published describing a novel developmental disorder caused by mutations in ADNP (ref. 27).

Figure 3 | Four novel genes with clustered mutations. The domains (blue), post-translational modifications, and mutation locations (red stars) are shown for four proteins with highly clustered de novo mutations in unrelated children with severe, undiagnosed developmental disorders. For the protein PCGF2, where all observed mutations are identical, photos are shown to highlight the facial similarities of patients carrying the same mutation.
phenotypes which could not be meaningfully linked to patient phenotypes. As many of our candidate developmental disorder genes carried heterogeneous loss-of-function variants (de novo mutations), it is to be expected that the severity of loss-of-function phenotypes in zebrafish may exceed that observed in our patient cohort. The genes with proven non-redundant developmental roles can reasonably be assigned higher priority for downstream functional investigations and genetic analyses.

Our systematic review of gene perturbation in multiple species sought both confirmatory and contradictory (for example, healthy homozygous knockout) evidence from other animal models for these 21 apparently developmentally important genes. We identified 16 genes with solely confirmatory data, often from multiple different organisms, none with solely contradictory data, two with both confirmatory and contradictory evidence, and three with no evidence either way (Supplementary Table 8).

In summary, our analyses validate a large-scale, genotype-driven strategy for novel developmental-disorder-linked gene discovery that is complementary to the traditional phenotype-driven strategy of studying patients with very similar presentations, and is particularly effective for discovering novel developmental disorders with highly variable or indistinct clinical presentations. Our meta-analysis with previously published developmental disorder studies increased power to detect variants in undiagnosed patients of functional mutations in genes predicted to operate recessively or by gain-of-function mechanisms will be lower than for haplosufficient genes; (3) the significant enrichment in undiagnosed patients of functional mutations in genes predicted to exhibit haploinsufficiency (Extended Data Fig. 6b); and (4) the strong enrichment for developmental phenotypes in the zebrafish knockdown screen.

Given our limited power to detect pathogenic mutations that act through dominant-negative or activating mechanisms, it was notable that in four of our novel genes (COL4A3BP, PPP2R1A, PPP2R5D and PCGF2) we observed identical de novo mutations in unrelated trios. Two hypotheses might explain this observation. First, that there is a vast number of different gain-of-function mutations, of which we are just scratching the surface in this study, or second, that these particular variants are enriched in our cohort owing to these mutations conferring a positive selective advantage in the germ line. Analysis of larger data sets will be required to assess these hypotheses, although they are not necessarily mutually exclusive.

These considerations of the limited power of even nationwide studies such as ours motivate the international sharing of minimal genotypic and phenotypic data, for example through the DECIPHER web portal (http://decipher.sanger.ac.uk), to provide diagnoses for patients who would otherwise remain undiagnosed. Plausibly pathogenic variants observed in undiagnosed patients in our study (de novo SNVs, indels and CNVs, and biallelic loss of function in genes not yet associated with disease) are shared through DECIPHER, and we encourage other, comparable studies to adopt a similar approach.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 July; accepted 4 December 2014.
Published online 24 December 2014.

1. OMIM. Online Mendelian Inheritance in Man http://omim.org (2014).
2. Cooper, G. M. et al. A copy number variation morbidity map of developmental delay. Nature Genet. 43, 838–846 (2011).
The Deciphering Developmental Disorders Study

T. W. Fitzgerald +14, S. S. Gerey +14, W. D. Jones +14, M. van Kogelenberg +14, D. A. King +14, J. McRae +14, K. A. Merry +14, V. Partanen +14, S. Al-Turki +14, K. Ambridge +14, D. M. Barrett +14, T. Berezinova +14, S. Clayton +14, E. L. Coomber +14, S. Gribble +14, P. Jones +14, N. Krishnapa +14, L. E. Mason +14, A. Middleton +14, R. Miller +14, E. Pigrimore +14, D. Rajan +14, A. Srinath +14, A. R. Tivey +14, M. Ahmed +14, N. Awake +14, R. Andrews +14, U. Anjum +14, A. Reche +14, R. Armstrong +14, M. Alabasuramanian +14, R. Banerjee +14, D. Barall +14, P. Batstone +14, D. Baty +14, M. Bick +14, J. Bingley +14, B. Blackman +14, S. Bhattacharaya +14, L. Boudin +14, D. Bourn +14, A. Brady +14, E. Braggin +14, E. Brewerton +14, L. Brunton +14, K. Brunstrom +14, S. J. Bumpstead +14, D. J. Burn +14, T. Burton +14, N. Canham +14, B. Castle +14, K. Chandler +14, S. Clasper +14, J. Clayston-Smyth +14, T. Cole +14, A. Collins +14, C. Connolly +14, N. Connelly +14, M. Cox +14, D. Coop +14, M. Crock +14, A. Deshpande +14, G. Devlin +14, A. Doxt +14, D. Dominick +14, C. Donnelly +14, D. Donnelly +14, A. Douglas +14, A. Duncan +14, J. Eason +14, E. Edkins +14, S. Ellard +14, P. Ellis +14, F. Elmslie +14, K. Evans +14, S. Everest +14, T. Ferndia +14, R. Fisher +14, F. Flinter +14, N. Foulds +14, A. Fryer +14, G. Gadd +14, S. Galloway +14, P. Galloway +14, J. Galloway +14, R. Goodship +14, D. Gouldey +14, E. Gray +14, P. Greene +14, L. Greenhalgh +14, L. Harrison +14, R. Hawkins +14, S. Hellens +14, A. Henderson +14, E. Hobson +14, S. Holder +14, G. Hollingsworth +14, T. Hornby +14, M. Humphreys +14, J. Hurst +14, S. Ingram +14, M. Irving +14, L. Jenkins +14, D. Johnson +14, D. Jones +14, D. Josifovski +14, S. Jones +14, K. Kaakinen +14, E. Kivuva +14, A. Kraus +14, D. Lachlan +14, K. Lachlan +14, W. Lamy +14, A. Lampé +14, C. Langman +14, M. Lees +14, D. Lin +14, G. Lowther +14, S. A. Lynch +14, A. Magee +14, E. Maher +14, M. Mansour +14, K. Marks +14, K. Martin +14, U. Maye +14, E. McCann +14, V. McConnell +14, M. McEntargart +14, R. McGowan +14, K. McKay +14, S. McKee +14, D. J. McMullan +14, S. McNealan +14, S. Mehta +14, K. Metcalfe +14, E. Miles +14, S. Mohammed +14, T. Montgomery +14, D. Moore +14, M. Morris +14, A. Morris +14, J. Morton +14, H. Mugalasy +14, V. Murday +14, L. Neville +14, R. Newbury-Ecob +14, A. Newman +14, R. O’Shea +14, J. O’Sullivan +14, L. O’Tregenna +14, M. Parker +14, P. Paton +14, P. Paton +14, J. Paton +14, S. Payne +14, J. Phillips +14, D. Pinto +14, D. Porteous +14, N. Pratt +14, K. Presscott +14, S. Price +14, A. Priddham +14, A. Proctor +14, H. Purcell +14, N. Ragge +14, J. Rankin +14, L. Raymond +14, D. Rice +14, L. Roberts +14, E. Roberts +14, G. Roberts +14, R. Roberts +14, S. Ross +14, S. S. Sancak +14, S. Sarchi +14, C. Scott +14, R. Scott +14, A. Selby +14, A. Seller +14, C. Sequeira +14, N. Shannon +14, S. Shairi +14, C. Shaw-Smith +14, E. Shearing +14, D. Shears +14, I. Simonic +14, D. Simpson +14, R. Singzon +14, Z. Skitz +14, A. Smith +14, B. Smith +14, K. Smith +14, S. Smithson +14, L. Sneddon +14, M. Smith +14, M. Squires +14, H. Stark +14, L. Stoodley +14, G. J. Swaminathan +14, E. Sweeney +14, K. Tatton-Brown +14, C. Taylor +14, R. Taylor +14, M. Tein +14, I. K. Tein +14, J. Thomson +14, A. Tokovka +14, B. Traynor +14, C. Turner +14, P. Turnerpeny +14, C. Tyrose +14, A. Vandersteen +14, P. Vasudevan +14, J. Vogt +14, E. Wakeling +14, D. Walker +14, J. Waters +14, A. Weber +14, D. Wellesley +14, M. Whiteford +14, S. Widsa +14, J. Wield +14, D. Williams +14, L. Wilson +14, Y. Yang +14, M. Yao +14, N. P. Carter +14, M. Parker +14, H. Firth +14, D. R. FitzPatrick +14, C. J. Wright +14, B. C. Jay +14, M. E. Hurles +14

1Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. 2Wessex Clinical Genetics Service, University Hospital Southampton, Park, Southampton General Hospital, Southampton SO16 6YD, UK. 3Wessex Regional Genetics Laboratory, Salisbury NHS Foundation Trust, Salisbury District Hospital, Oastbrook Road, Salisbury, Wiltshire SP2 8BJ, UK. 4Faculty of Medicine, University of Southampton, Southampton SO16 6YD, UK. 5South West Thames Regional Genetics Centre, St George’s Healthcare NHS Trust, St George’s, University of London, Cranmer Terrace, London SW17 0RE, UK. 6School of Medical Genetics, University of Wales, Heath Park, Cardiff CF14 4XW, UK. 7Department of Clinical Genetics, Block 12, Glan Clwyd Hospital, Rhyl, Denbighshire LL18 5JU, UK. 8East Anglian Medical Services Group, Box 134, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK. 9Sheffield Regional Genetics Services, Sheffield Children’s NHS Trust, Western Bank, Sheffield S10 2TJ, UK. 10North of Scotland Regional Genetics Service, NHS Grampian, Department of Medical Genetics Medical School, Aberdeen AB25 2ZD, UK. 11East of Scotland Regional Genetics Service, Human Genetics Unit, Pathology Department, NHS Tayside, Ninewells Hospital, Dundee DD1 9SY, UK. 12Yorkshire Regional Genetics Service, Leeds Teaching Hospitals NHS Trust, Department of Clinical Genetics, Chapel Allerton Hospital, Chapelton Road, Leeds LS5 6QA, UK. 13West North Thames Regional Genetics Centre, North West London Hospitals NHS Trust, The Kennedy Galton Centre, Northwick Park and St Mark’s NHS Trust Watford Road, Harrow HA1 3UJ, UK. 14Oxford Regional Genetics Service, Oxford Radcliffe Hospitals NHS Trust, The Churchill Old Road, Oxford OX3 7UL, UK. 15West Midlands Regional Genetics Service, Birmingham Women’s NHS Foundation Trust, Birmingham Women’s Hospital, Edgbaston, Birmingham B15 2TG, UK. 16Northern Genetics Service, Newcastle upon Tyne Hospitals NHS Foundation Trust, Institute of Human Genetics, International Centre for Life, Central Park, Newcastle upon Tyne NE2 4XL, UK. 17Peninsula Clinical Genetics Service, Plymouth Hospitals NHS Trust, University Hospital PL2 6GJ, UK. 18North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Foundation Trust, Great Ormond Street, Great Ormond Street, London WC1N 3JH, UK. 19Manchester Centre for Genomic Medicine, St Mary’s Hospital, Central Manchester University Hospitals NHS Trust, Manchester Children’s Academic Health Science Centre, Manchester M13 9WL.
UK. 20 South East Thames Regional Genetics Centre, Guy’s and St Thomas’ NHS Foundation Trust, Guy’s Hospital, Great Maze Pond, London SE1 9RT, UK.
21 Leicestershire Genetics Centre, University Hospitals of Leicester NHS Trust, Leicester Royal Infirmary (NHS Trust), Leicester LE1 5WW, UK. 22 Nottingham Regional Genetics Service, City Hospital Campus, Nottingham University Hospitals NHS Trust, The Gables, Hucknall Road, Nottingham NG5 1PB, UK. 23 Northern Ireland Regional Genetics Centre, Belfast Health and Social Care Trust, Belfast City Hospital, Lisburn Road, Belfast BT9 7AB, UK. 24 West of Scotland Regional Genetics Service, NHS Greater Glasgow and Clyde, Institute Of Medical Genetics, Yorkhill Hospital, Glasgow G3 8SJ, UK. 25 College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK. 26 Merseyside and Cheshire Genetics Service, Liverpool Women’s NHS Foundation Trust, Department of Clinical Genetics, Royal Liverpool Children’s Hospital Alder Hey, Eaton Road, Liverpool L12 2AP, UK. 27 MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. 28 Bristol Genetics Service (Avon, Somerset, Glos and West Wilts), University Hospitals Bristol NHS Foundation Trust, St Michael’s Hospital, St Michael’s Hill, Bristol BS2 8DT, UK. 29 National Centre for Medical Genetics, Our Lady’s Children’s Hospital, Crumlin, Dublin 12, Ireland. 30 School of Molecular, Genetic and Population Health Sciences, University of Edinburgh Medical School, Teviot Place, Edinburgh EH8 9AG, UK. 31 University of Edinburgh, Institute of Genetics & Molecular Medicine, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK. 32 School of Medicine, Dundee University, Mackenzie Building, Kirsty Semple Way, Ninewells Hospital and Medical School, Dundee DD2 4RB, UK. 33 The Ethox Centre, Nuffield Department of Population Health, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK.

*These authors contributed equally to this work.
Extended Data Figure 1 | Characteristics of the families. a, Gestation-adjusted decimal age (years) at last clinical assessment. The histogram shows the distribution of the gestation-adjusted decimal age at last clinical assessment across the 1,133 probands. The dashed red line shows the median age. b, Frequency of human phenotype ontology (HPO) term usage. Bar plot showing, for each used HPO term, the number of times it was observed across the 1,133 proband patient records. c, Projection PCA plot of the 1,133 probands. PCA plot of 1,133 DDD probands projected onto a PCA analysis using four different HapMap populations from the 1000 genomes project. Black, African; red, European; green, east Asian; blue, south Asian; and the 1,133 DDD probands are represented by orange triangles. d, Self-declared and genetically defined consanguinity. Overlaid histogram showing the distribution of kinship coefficients from KING comparing parental samples for each trio. Green, trios where consanguinity was not entered in the patient record on DECIPHER; red, trios consanguinity was declared in the patient record on DECIPHER.
Extended Data Figure 2 | Number of validated de novo SNVs and indels per proband. Bar plot showing the distribution of the observed number of validated SNVs and indels per proband sample, and the expected distribution assuming a Poisson distribution with the same mean as the observed distribution.
Extended Data Figure 3 | Number of diagnoses per gene. Histogram showing the number of diagnoses per gene for genes with at least two diagnoses from different proband samples.
Extended Data Figure 4 | Burden of large CNVs in 1,133 DDD proband samples. Plot comparing the frequency of rare CNVs in three sample groups against CNV size. The y axis is on a log scale. Red, DDD probands who have not had previous microarray-based genetic testing; purple, DDD probands who have had negative previous microarray-based genetic testing; green, DDD controls.
Extended Data Figure 5 | Expected and observed numbers of de novo mutations. The expected and observed numbers of mutations of different functional consequences in three mutually exclusive sets of genes are shown, along with the $P$ value from an assessment of a statistical excess of observed mutations. The three classes of genes are described in the main text.
Extended Data Figure 6 | Haploinsufficiency analyses. a. Saturation analysis for detecting haploinsufficient developmental-disorder-linked genes. A box plot showing the distribution of statistical power to detect a significant enrichment of loss-of-function mutations across 18,272 genes in the genome, for different numbers of trios studied, from 1,000 trios to 12,000 trios. Line within the box shows the median, box shows the interquartile range and the whiskers show the most extreme values within 1.5 times the interquartile range from the box. b. Distribution of haploinsufficiency scores in selected sets of de novo mutations. Violin plot of haploinsufficiency scores in five sets of de novo mutations. Silent, all synonymous mutations; diagnostic, mutations in known developmental-disorder-linked genes in diagnosed individuals; undiagnosed_Func, all functional mutations in undiagnosed individuals; undiagnosed_LoF, all loss-of-function mutations in undiagnosed individuals; undiagnosed_recur, mutations in genes with recurrent functional mutations in undiagnosed individuals. P values for a Mann–Whitney U-test comparing each of the latter four distributions to that observed for the silent (synonymous) variants are plotted at the top of each violin. Dot indicates the median, box is interquartile range and whiskers are the most extreme values within 1.5 times the interquartile range from the box.
Extended Data Table 1 | Novel genes with suggestive evidence for a role in developmental disorder

| Evidence                      | Gene  | de novos DDD (Missense, LoF) | de novos Meta (Missense, LoF) | P Value       | Test  | Mutation Clustering | Predicted Haploinsufficiency |
|-------------------------------|-------|------------------------------|-------------------------------|---------------|-------|--------------------|-------------------------------|
| De novo enrichment +          | NAA15 | 1 (0,1)                      | 3 (0,3)                       | 1.64E-06      | Meta  | No                 | 7.5%                          |
| additional evidence           | ZBTB20| 3 (1,2)                      | 3 (1,2)                       | 4.84E-06      | DDD   | No                 | 0.2%                          |
|                               | NAA10 | 2 (2,0)                      | 3 (3,0)                       | 8.28E-06      | Meta  | No                 | 34.1%                         |
|                               | TRIP12| 3 (1,2)                      | 4 (2,2)                       | 2.13E-05      | Meta  | No                 | 3.8%                          |
|                               | USP9X | 3 (1,2)                      | 3 (1,2)                       | 5.14E-05      | DDD   | No                 | 3.8%                          |
|                               | KAT6A | 2 (0,2)                      | 2 (0,2)                       | 7.91E-05      | DDD   | No                 | 19.0%                         |

Six genes with suggestive evidence to be novel developmental-disorder-linked genes. The number of unrelated patients with independent functional or loss-of-function mutations in the DDD cohort or the wider meta-analysis data set including DDD patients is listed. The P value reported is the minimum P value from the testing of the DDD data set and the meta-analysis data set. The data set that gave this minimal P value is also reported. Mutations are considered to be clustered if the P value of clustering of functional SNVs is less than 0.01. Predicted haploinsufficiency is reported as a percentile of all genes in the genome, with 0% being highly likely to be haploinsufficient and 100% very unlikely to be haploinsufficient, based on the prediction score described in ref. 26 updated to enable predictions for a higher fraction of genes in the genome. NAA10 is already known to cause an X-linked recessive developmental disorder in males, but here we identified missense mutations in females, suggesting a different, X-linked dominant, disorder.
Extended Data Table 2 | Biallelic loss of function and damaging functional variants

| Biallelic Variant Types                  | Untransmitted Diploptypes (n=1080) | Likely Dominant Probands (n=270) | Other Probands (n=810) |
|------------------------------------------|-------------------------------------|----------------------------------|------------------------|
| LoF/LoF (Genome-wide)                   | 110                                 | 17                               | 86                     |
| LoF/Dam (Genome-wide)                   | 87                                  | 21                               | 71                     |
| Dam/Dam (Genome-wide)                   | 312                                 | 90                               | 264                    |
| LoF/LoF (DDG2P Biallelic)               | 1                                   | 1                                | 3                      |
| LoF/Dam (DDG2P Biallelic)               | 2                                   | 0                                | 6                      |
| Dam/Dam (DDG2P Biallelic)               | 26                                  | 7                                | 25                     |

Rare (MAF <5%) biallelic loss of function and damaging functional variants in uninherited diploptypes and probands. ‘Likely dominant probands’ refers to probands with a reported de novo mutation or affected parents, and ‘other probands’ refers to all remaining probands. ‘DDG2P biallelic’ refers to confirmed and probable DDG2P genes with a biallelic mode of inheritance. See Supplementary Methods for details of variant processing.
Extended Data Table 3 | Zebrafish modelling identifies 21 developmentally important candidate genes

| Gene   | # patients | Variant         | Patient phenotypes                                      | Phenotypic concordance | Relevant knockdown phenotypes          |
|--------|------------|-----------------|--------------------------------------------------------|------------------------|----------------------------------------|
| BTBD9  | 2/1        | Biallelic LoF/De novo Missense | Seizures, microcephaly, hypertonia                     | Strong                 | Reduced head size, brain volume        |
| CHD3   | 1/2        | De novo LoF/Missense                | CNS and craniofacial defects                           | Strong                 | Abnormal head shape                    |
| DDX3X  | 1/3        | De novo LoF/Missense                | Moderately short stature, microcephaly, CNS defects    | Strong                 | Reduced head size, brain volume        |
| ETF1   | 1          | De novo LoF                        | CNS and craniofacial defects, seizures, microcephaly, hypertelorism | Strong                 | Reduced head size, brain volume        |
| FRYL   | 1          | De novo LoF                        | Short stature, craniofacial and cardiac defects        | Strong                 | Cardiac defects, reduced axis length   |
| PKN2   | 1          | De novo Missense                   | CNS, cardiac, ear, and craniofacial defects, growth retardation | Strong                 | Cardiac, craniofacial cartilage, and growth defects |
| PSMD3  | 1          | De novo Missense                   | Microcephaly, muscular hypotonia, seizures, growth abnormality | Strong                 | Reduced head size and neural defects   |
| SCSN   | 1          | Biallelic LoF                      | Seizures, microcephaly, CNS defects                   | Strong                 | Reduced head size, brain volume        |
| SETDS  | 1          | De novo LoF                        | Seizures, CNS and cardiac defects, poor motor coordination | Strong                 | Reduced head size, cardiac defects, abnormal locomotion |
| THNSL2 | 2          | Biallelic LoF                      | Microcephaly, CNS and ear defects                     | Strong                 | Reduced head size, brain volume, neural defects |
| ZRANB1 | 2          | De novo Missense                   | Microcephaly, muscle defects, seizures                | Strong                 | Reduced head size and neural defects    |
| DPEP2  | 1          | Biallelic LoF                      | CNS defects, growth retardation                        | Moderate               | Growth reduction                        |
| PSD2   | 1          | De novo LoF                        | CNS defects, hypotonia, seizures                       | Moderate               | Abnormal musculature, CNS and locomotion |
| SAP130 | 1          | De novo LoF                        | Short stature, hypotonia, hypotelorism                | Moderate               | Abnormal locomotion                    |
| CNOT1  | 1/1        | De novo LoF/Missense               | Short stature, cardiac, CNS, ear and craniofacial defects | Weak                   | Multisystem                            |
| DTWD2  | 1          | De novo LoF                        | CNS defects, seizures                                 | Weak                   | Multisystem                            |
| ILVBL  | 1          | De novo LoF                        | CNS and craniofacial defects                          | Weak                   | Multisystem, with otic and growth defects |
| NONO   | 1          | De novo LoF                        | CNS and ear defects, hypotonia, growth retardation     | Weak                   | Multisystem                            |
| POGZ   | 2          | De novo LoF                        | CNS and ear defects, hypotonia, coloboma              | Weak                   | Multisystem                            |
| SMARCD1| 1/1        | De novo LoF/Missense               | CNS defects, hypotonia                                | Weak                   | Multisystem                            |
| WWC1   | 1          | De novo Missense                   | CNS defects, hypotelorism                             | None                   | None                                   |

This table summarizes the 21 genes for which knockdown results in developmental phenotypes in zebrafish. The ‘# patients’ column indicates how many patients were identified as carrying variants in these genes. Split numbers indicate the breakdown of variant types (for example, for BTBD9, 2/1 is two biallelic loss of function and one de novo missense carrying patients). A summary of the patient phenotypes is listed, as well as the relevant phenotypes observed in zebrafish knockdown experiments. Phenotypic concordance categories indicate the degree of overlap between the zebrafish phenotyping and the patient phenotypes. Weak concordance typically is the result of severe, multisystem phenotypes in zebrafish. See Supplementary Information for more detailed phenotype information.