**Human myelomeningocele risk and ultra-rare deleterious variants in genes associated with cilium, WNT-signaling, ECM, cytoskeleton and cell migration**

K. S. Au, L. Hebert, P. Hillman, C. Baker, M. R. Brown, D.-K. Kim, K. Soldano, M. Garrett, A. Ashley-Koch, S. Lee, J. Gleeson, J. E. Hixson, A. C. Morrison & H. Northrup

Myelomeningocele (MMC) affects one in 1000 newborns annually worldwide and each surviving child faces tremendous lifetime medical and caregiving burdens. Both genetic and environmental factors contribute to disease risk but the mechanism is unclear. This study examined 506 MMC subjects for ultra-rare deleterious variants (URDVs, absent in gnomAD v2.1.1 controls that have Combined Annotation Dependent Depletion score ≥ 20) in candidate genes either known to cause abnormal neural tube closure in animals or previously associated with human MMC in the current study cohort. Approximately 70% of the study subjects carried one to nine URDVs among 302 candidate genes. Half of the study subjects carried heterozygous URDVs in multiple genes involved in the structure and/or function of cilium, cytoskeleton, extracellular matrix, WNT signaling, and/or cell migration. Another 20% of the study subjects carried heterozygous URDVs in candidate genes associated with gene transcription regulation, folate metabolism, or glucose metabolism. Presence of URDVs in the candidate genes involving these biological function groups may elevate the risk of developing myelomeningocele in the study cohort.

Successful neural tube formation relies on a series of orchestrated biological processes to facilitate convergent extension of neural ectodermal cells (NE) on the neural plate. NE cells at the midline and the dorsolateral regions form the medial hinge point (MHP) and the dorsolateral hinge point (DLHP) through mechanisms including apical construction and localized cell proliferation, and proliferation of non-neural ectodermal cells (NNE) that facilitate closing of the neural tube (Fig. 1). These biological processes involve sensing signals in the extracellular matrix by signal receptors on the cell surface and sub-organelles (e.g. mecanochemical sensors on primary cilia) which subsequently trigger remodeling of cytoskeletal structures and orchestrate directional migration of NE and NNE cells. The neural tube can fail to close due to impaired convergent extension, MHP or DLHP formation, structural integrity of NE, or midline fusion of NE or NNE. Maintaining NNE integrity and controlling epithelial-to-mesenchymal-transition of neural crest cells, and balancing proliferation of hindgut cells (of endodermal origin) may also play important roles in the development of neural tube. Synchrony of all these processes together transform the neural plate into neural folds that meet and merge to form the neural tube. Factors interrupting the structure and/or function and spatial temporal integrity of these synchronous processes could result in failure of neural tube closure.

More than 400 genes have been identified in animal models that are required to maintain normal neural tube closure. Many of the neural tube defect (NTD) animal model
genes constitute structural and/or functional components of cilium, extracellular matrix, the WNT signaling pathway, actomyosin cytoskeleton remodeling and cell migration.

An increasing number of heterozygous deleterious genetic variants in genes regulating planar cell polarity (PCP) have been revealed in human NTD cases and suggest that multi-genic heterozygous deleterious variants may contribute to NTD development. Knowledge is lacking regarding the extent of deleterious variants in genes involving various biological processes contributing to human myelomeningocele (MMC). It is possible that defective genes not known to interact with PCP genes in combination with defective PCP genes together can confer risk of MMC. Genes of interest for this study included variants of 551 candidate genes (537 mouse genes and 14 human NTD associated genes) previously shown to cause NTDs in mouse embryos or previously associated with human NTD (see Supplementary Table 1). In a complex trait study, the effect size of genetic variants has been shown to be inversely related to the frequency of the alternate allele of variants, and new and private variants could constitute an important reservoir of disease risk alleles. Single nucleotide variants (SNVs) in 551 candidate genes that were absent in non-Finnish European (NFE) or Ad-mixed American (AMR) in gnomAD v2.1.1 (controls) and which have top 1% deleteriousness (Combined Annotation Dependent Depletion phred score, C-score ≥ 20) were catalogued and defined as ultra-rare deleterious variants (URDVs) for analysis in the

**Figure 1.** Neural tube formation involves a series of biological processes including mechanochemical signal sensing from the environment to orchestrate cell proliferation, cytoskeleton remodeling, and migration of neural and non-neural ectodermal cells. (a) Neural plate cells undergo convergent-extension to facilitate posterior-anterior elongation of the neural tube. Neuroectodermal cells (NE) migrate to the midline in response to extracellular signals received by cilia in NE cells to establish and maintain polarity. (b) Midline NE cells undergo apical constriction via a biological process involving actomyosin cytoskeleton remodeling to form the midline hinge point (MHP). NE cell and non-neuroectodermal (NNE) cells continue proliferation and migration during the neural tube closing process upon sensing signals in the extracellular matrix (ECM). NE cells at the dorsolateral hinge point (DLHP) undergo proliferation, migration dorsally, and reshaping to form the DLHP bringing the NE/NNE edges at the dorsal midline together. Advancing migration and proliferation of NNE and mesenchymal cells facilitate ectodermal layer extension when dorsal midline cells' protrusions meet and join at the closing point. The process also brings NE cells together to establish cell junctions and close the NT.
study. The majority of the homozygous defective mouse genes caused open neural tubes in the cranial regions whereas heterozygous mice had normal neural tube development. Interestingly, some mice with a subset of these genes in the digenic heterozygote condition developed spina bifida, suggesting that oligogenic heterozygosity is a possible mechanism for spina bifida development. Recent human spina bifida studies revealed cases carrying only heterozygous damaging variants of PCP genes and examples of subjects with digenic heterozygote rare variants in some PCP genes were observed. In addition, some cases had damaging variants in two PCP genes each inherited from one of the parents. Oligogenic inheritance is a demonstrated disease mechanism in many human diseases including the birth defect with holoprosencephaly. Representation of oligogenic heterozygous URDVs in two or more NTD candidate genes found in MMC subjects in the study will be reported. Ontology analysis suggested URDV-containing NTD candidate genes were enriched for the same ontology group or in multiple categories including cilium, cytoskeleton, extracellular matrix, WNT signaling, and cell migration. The potential contribution of the observed results to MMC development in humans will be discussed.

Results

Global landscape of single nucleotide variants of MMC exomes. Three European (EA) and two Mexican (MA) MMC subjects had variant counts less than 80% of the median variant counts of the study subjects and were removed from further analysis, leaving 254 EA and 252 MA MMC exomes. A total of 597,401 high confidence SNVs that passed filtering steps described in the Materials and Method section were selected for further analysis. Using dbNSFP v4.0 to annotate high quality SNVs, 112,369 variants predicted to be functionally significant in 16,135 genes were selected for further analysis that included 109,727 missense, 1814 stop_gained, 84 stop_lost and 771 splice site variants. Non-coding SNVs and synonymous SNVs were not analyzed in this study.

Using AMR and NFE datasets from gnomAD v2.1.1 as controls, approximately 15.1% and 13.2% of the SNVs in EA MMC and MA MMC respectively were ultra-rare variants (URVs) with alternate allele frequency (aAF) in the gnomAD v2.1.1 (controls) equal to zero or with an absence of the known variant. URVs found only in MMC subjects may constitute an important enriched pool of risk alleles for further analysis. An average of 40 URVs per subject were identified. Over 98% of URVs occurred only once. Approximately 7400 variants is inversely proportionate to the alternate allele frequencies (aAF) in each ethnic group (see Supplementary Table 2). These variants were defined as ultra-rare functional deleterious SNVs (URDVs) and used for further analysis. Each subject had approximately 29 URDVs with range between 9 and 90.

Eighty-eight subjects were re-sequenced using the Illumina NGS platforms. Of the 30,789 SNVs in 33 EA and 37,924 SNVs in 55 MA identified from the Ion Proton platform in this study, over 98.5% were also identified by the Illumina NGS platforms.

Landscape of URVs in NTD candidate genes in human MMC exomes. There are several hundred genes known to cause NTDs in mice under monogenic and oligogenic conditions. We sought insight through examining SNVs in 551 NTD candidate genes (see Supplementary Table 1) known to cause abnormal neural tube closure in animal models (537 genes), and 14 genes associated with human MMC in our study cohort. A total of 2134 SNVs in 440 genes were identified from EA MMC exomes and 2496 SNVs in 447 genes were identified from MA MMC exomes.

Ultra-rare functional deleterious SNVs (URDVs) in candidate genes. Many studies have suggested the impact of variants is inversely proportionate to the alternate allele frequencies (aAF). Using aAF of variants in NFE/AMR gnomAD (controls) exomes as a filter, all 506 MMC subjects carried 31–69 rare deleterious variants (RDVs, aAF<0.01) in NTD candidate genes whereas the number of URDVs per subject varied between zero and eight (see Fig. 2 & Supplementary Table 3). With increased findings of novel and very rare variants in PCP genes associated with human NTDs in many previous studies, we chose to focus on the novel URDVs, which are more likely to have higher effect size, for subsequent analyses.

Among 254 EA MMC subjects there were 307 URDVs in 202 genes. The range of URDVs per gene found in EA MMC was zero to six. Among 252 MA MMC subjects, there were 331 URDVs found in 203 genes. An average of 1.63 URDVs in a URDV-containing NTD candidate gene was identified in MA MMC group. The range of URDVs per gene found in MA MMC subjects was zero to nine. Over 97% of URDVs had an allele count (AC) of one and a few had two to three. The majority (~95%) of URDVs were missense changes while the remaining ~3.4% were stop_gained or splice site changes (1.3% and 2.4% in EA and MA groups respectively). Around 70% of MMC subjects in both ethnic groups had at least one URDV in an NTD candidate gene, with the highest URDV counts being eight and five in EA MMC and MA MMC subjects, respectively (Fig. 2). There were 78 EA MMC subjects and 67 MA MMC subjects who had no URDVs in NTD candidate genes.

URDV-containing NTD candidate genes were either common to both ethnicities or found uniquely in EA MMC and MA MMC exomes. There were 99 genes unique to EA MMC, 100 genes unique to MA MMC and 103 genes were common to both ethnic groups. Examining EA MMC and MA MMC respectively, most (66.7% and 63.4%) of the genes had one URDV; 21.3% and 22.1% had two; 10.9% and 7.4% had three and the remaining few had four to nine URDVs. For the entire study cohort, 636 URDVs with 603 missense, 21 stop_gained and 12 splice site changes in 302 NTD candidate genes were discovered (Supplementary Table 1). These URDVs were manually verified using the online Integrative Genomics Viewer (IGV) Web App (https://igv.org/app). Seven subjects carried two heterozygous URDVs in their candidate genes (Supplementary Table 3). Visual inspection using IGV revealed the two heterozygous URDVs were in cis positions in the candidate genes (i.e. FREM2, POLE2, RPRGIP1L, BMP2, and CGN) of five subjects respectively. One subject (i.e. C35-874, Supplementary
Table 3) had two heterozygous URDVs in trans position in the *KIF7* gene. The remaining subject (i.e. F57-264, Supplementary Table 3) carried two URDVs 1400 bp apart in two different exons of *PNPLA6* therefore, we were unable to determine whether the URDVs were in cis or in trans.

The number of URDVs identified was not directly proportionate to the length of coding sequence (CDS) of genes in the two subject groups (Supplementary Table 3). Genes with longer transcript length (CDS length > 10 Kb) had a higher number of URDVs as expected (e.g. *FREM2*, *HSPG2*, *APOB*, *DYN2H1* and *HTT*) although their URDVs/Kb CDS was less than one. Overall, 96 genes had URDV densities of over one per Kb of CDS and the 64 genes had ≥ 1.25 URDV/Kb CDS (Fig. 3, Supplementary Table 3). These 96 genes harboring 272 URDV accounted for over 40% of the total 636 URDV identified in 302 candidate genes. The median size of the CDS for the 96 genes was 1.39 Kb.

Nine genes in MA MMC subjects had a higher number of URDVs, ranging from four to nine, with the most in *FREM2*. In EA MMC subjects, eight genes had four to six URDVs with six in *APOB* and in *FREM2*. Three genes (*EP300*, *HSPG2*, and *FREM2*) found in both EA and MA MMC subjects had more than ten URDVs with *FREM2* and *EP300* had an URDV density around 1.5.

**Figure 2.** Distributions of URDVs Identified by WES. (a) Chart shows count of URDVs in NTD candidate gene(s) identified per exome of MM subjects varied from 0 to 8. In this study, the number of EA MMC subjects is 254 and the number of MA MMC subjects is 252. Approximately 25–30% subjects had no URDVs in NTD candidate genes, around 35% had one URDV, and the remaining had more than one. (b) Shows counts of URDVs discovered per NTD candidate gene ranged from one to nine with nearly 65% genes containing only one. Total number of URDV-containing NTD candidate genes found in EA and MA MMC subjects are 202 and 203 respectively.
Figure 3. URDV density per Kb coding sequence of NTD candidate genes. The majority of 302 NTD candidate genes show URDV/KbCDS density less than one. The x-axis showed the gene names ordered by the URDV/KbCDS from highest (4.1 for CTNNBP1) to lowest (1.28, CYP26C1). Dashed line (---) showed the URDV/KbCDS ratio of the gene and dotted line (………) showed the length of CDS in Kb. Count of URDV in EA showed in blue bar and MA showed in red bar. Approximately half of the 64 genes had URDVs found in both EA and MA. One-third of the 64 genes have ≥ 2 URDVs/Kbp CDS.
Ontology Enrichment of NTD candidate genes with URDVs in MMC exomes. Among the 551 NTD candidate genes, enrichment of components in cilium, cytoskeleton, extracellular matrix and cell migration/adhesion, and WNT signaling were suggested (Supplementary Table 4). Using the ToppCluster online program, genes between EA and MA groups consisting of URDVs were compared to the group of genes lacking URDVs using Bonferroni correction for multiple testing. Analysis results showed genes in EA and MA groups had higher folds of enrichment in ontology classes than those of the overall 551 candidate genes. However, there was very distinct overrepresentation of subclasses of ontology for URDV-containing candidate genes in the subjects and these subclasses were absent among the candidate genes without URDVs (Fig. 4 and Supplementary Table 4).

Enrichment of cellular component genes containing URDVs was seen mostly among ciliary components and intracellular transport, actin filament, and microtubule cytoskeleton organization in both ethnic groups (Fig. 4a, and Supplementary Table 4). URDV-containing genes in EA MMC subjects were associated with components found in neuronal cell body, perinuclear region of cytoplasm, ciliary transition fiber, and WNT/SHH signaling. URDV-containing genes in MA MMC subjects were associated with components of ciliary basal body actin-based cytoskeleton and cell projection and cell–cell junctions. Differences in cellular component genes with URDV between EA and MA subjects were observed.

Molecular function analysis showed enrichment of URDV-containing candidate genes with molecular functions in transcription modulation, beta-catenin binding, cell adhesion, and cytoskeleton binding in both EA and MA MMC exomes (Fig. 4b, and Supplementary Table 4). EA MMC cases had enriched candidate genes related to hedgehog family protein and WNT signaling molecule binding. MA MMC cases had enriched candidate genes involved in DNA binding with transcription factors and nuclear hormone receptors. Together, nearly half of the MMC subjects in the study cohort had one or more URDVs in the candidate genes belonging to the ontological groups associated with ciliary structures and functions.

Both EA and MA MMC exomes consistently had URDV-containing genes disproportionately represented in the Hedgehog/WNT signaling and cancer signaling pathways (Fig. 4c, and Supplementary Table 4). Of note, EA MMC exomes uniquely enriched with URDVs in genes from WNT/PCP, SHH and TGFβ signaling pathways. MA MMC exomes were uniquely enriched with URDVs in genes related to thyroid hormone signaling, NOTCH signaling, and anchoring ciliary basal body.

Most of the NTD candidate genes involve biological processes of embryo development of organs particularly for the central nervous system and other organ systems (Supplementary Table 4). Ontology terms for biological processes are also consistent with those described in the cellular components and molecular functions sections (Fig. 4d).

Ciliary structure and functions are closely associated with GO terms for cytoskeleton/microtubules, extracellular matrix, WNT signaling, and migration of cells and cellular components. To further dissect the properties of URDV-containing NTD candidate genes and to examine the URDV distribution in MMC exomes, we performed analysis of genes assigned to these five GO terms (Supplementary Table 5). Candidate genes in each of these ontological groups consisted a range of 12–23.2% of the total URDVs (Table 1). Together, nearly half of the MMC subjects in the study cohort have one or more URDVs in the candidate genes belonging to the five ontological groups. There were 113 and 110 NTD candidate genes with URDVs respectively in the EA and MA subject groups that were associated with functions for cilium, ECM, cytoskeleton and WNT signaling and cell migration. Of these genes, some belong in only one GO term and many in two to four GO terms with cross-interacting roles (Fig. 5, and Supplementary Table 6). Of 364 subjects in the study cohort who bore at least one URDV in an NTD candidate gene, nearly 75% had URDVs affecting genes associated with cilium, WNT signaling, ECM, cytoskeleton and cell migration remodeling suggesting they may confer increased genetic risk to MMC development. The remaining 25% of MMC subjects had URDV-containing NTD candidate genes that were associated with other gene functions such as transcription modulation, folate one carbon metabolism network and glucose oxidative stress not related to the ontologies examined in this study.

Subjects had multiple URDV-containing genes affecting one or multiple ontologies. Recently published articles demonstrated the genetic contribution in individuals affected with NTDs that were potentially digenic heterozygote with de novo and/or rare deleterious variants inherited from each parent. In this study, nearly 30% of MMC subjects had two or more URDVs in different NTD candidate genes belonging to the same and/or different ontologies. Overall, 85 EA MMC and 91 MA MMC exomes had multiple URDV-containing genes classified with different ontologies (Supplementary Table 3). Approximately 40% of these subjects had URDV-containing NTD candidate genes representing various combinations of the five ontologies presented above.

Oligogenic heterozygote URDVs involving the same ontology. A subset of MMC subjects including 25 EA and 21 MA were carrying two to three URDV-containing candidate genes that belong to the same ontology group (Table 2). Nine EA MMC and four MA MMC subjects had URDVs found in two candidate genes associated with cilia structure/function (Table 2). Of note, many cilium gene products also play roles in one or more of the other four ontologies examined here. Two EA MMC and five MA MMC subjects had URDVs in two genes with functions associated with WNT signaling (Table 2). Eleven EA MMC and nine MA MMC subjects had URDVs found in two genes with functions associated with cytoskeleton structures and functions (Table 2). Three EA MMC and one MA MMC subjects had URDVs found in two or three genes with functions associated with ECM structure and functions genes (Table 2). Finally, seven EA MMC and ten MA MMC subjects had URDVs found in two or more candidate genes regulating cell migration (Table 2). Fifteen EA MMC subjects and 20 MA MMC subjects were carrying multiple URDV-containing candidate genes in different ontological groups.
Figure 4. Results of ontology and pathway enrichment analysis of candidate genes with and without URDVs identified. Enrichment analysis of (a) cellular components, (b) molecular functions, (c) pathways, and (d) biological processes, were performed using online tools ToppFun and ToppCluster. Count of URDV-containing genes in ontology class for EA (blue) and MA (red) MMC subjects and genes without URDVs (green) were shown with bar charts. Enrichment folds were represented with dashed lines for URDV-containing genes in EA (blue), URDV-containing genes in MA (red) and genes without URDVs (green). Enrichment comparisons passed Bonferroni correction. Detail descriptions of ontology, gene names, fold enrichment, and Bonferroni corrected P values can be found in Supplementary Table 4.
URDV-containing genes and human neural tube expression and mouse phenotypes. The expression profiles of genes for four or more human neural tubes at Carnegie Stages (CS) 12 and 13 by SAGE were available for use to annotate NTD candidate genes with URDVs22. Expression of nearly 40% of the 302 URDV-containing NTD candidate genes was consistently detectable in CS12 and/or CS13 human neural tubes compared to less than 15% of 249 genes with no URDV identified (Supplementary Table 2). A Chi-square test comparing NTD candidate genes with and without URDVs showed significantly more URDV-containing genes expressed in human CS12 and CS13 (p < 0.0001, Supplementary Table 2). In addition, over 80% of the 302 URDV-containing NTD candidate genes caused embryonic lethality on or before neural tube closure in knockout mice of these genes9.

Discussion
This study revealed that 70% of the 506 MMC subjects consisting of the two ethnic groups with the highest prevalence of myelomeningocele in North America carry ultra-rare deleterious variants (URDVs) in 302 genes previously demonstrated to cause NTD phenotypes in animal models or associated with human NTDs (Supplementary Table 2). Ontology enrichment analyses showed around 50% of subjects in the cohort have one or more URDV-containing NTD candidate genes potentially impairing the structure and/or function of one or more ontological groups including cilium, SHH and WNT signaling, remodeling ECM and cytoskeleton, and cell migration. Normal structure and function of these genes are necessary for successful closure of the neural tube in animals1,3. The study results shown here may have a high impact on our understanding of the genetic mechanisms of human MMC.

One-fifth of MMC subjects in the study had URDVs in NTD candidate genes affecting cilium structure and/or function, suggesting human myelomeningocele risk is very likely associated with ciliopathy. The extent of URDV-containing ciliopathy genes associated with human NTD cases is not known. A review article by Vogel et al. (2012) noted that some human syndromes such as Meckel-Gruber syndrome and Joubert Syndrome that present with NTDs were associated with ciliopathies and recommended genetic screening of ciliopathy genes.
normal function of core PCP maintenance genes (e.g. PTK7/VANGL1, SMURF2/LRP6) strongly support the strategy of screening PCP genes to discover risk alleles contributing to human NTD cases.9

The PCP pathway modulates cellular functions such as ciliogenesis, actomyosin cytoskeleton and microtubules remodeling that are critical to neural tube formation.13–16 Normal ciliogenesis is important for non-canonical WNT/PCP signaling (i.e. CELSR1/LRP6), Hedgehog off state (i.e. DYNC2H1‑HTT), and dynein intermediate chain binding (i.e. DYNC2H1‑HTT). It is likely that two genes within the same sub-functional group could contribute to digenic impairment of ciliary structure and/or functions critical to neural tube closure increasing the risk of MMC in these cases. For human NTD cases.35 Both NE and NNE involve the primary cilia playing roles in transducing extracellular signals to regulate cell growth, proliferation, directional migration and adhesion vital to neural tube development.1,2,24–26 Many mutant mouse models with loss of function in the cilia-associated genes developed NTDs.36 A high proportion of copy number variations in cilia genes have also been found in subjects affected by NTDs.83 In this study, heterozygous URDVs identified in B9D1, CC2D2A, INPP5E, KIAA0586, KIF7, MKS1, NPHP3, PIBF1, RPGRIP1L, and TMEM67 belong to ciliopathy genes known to contribute to autosomal recessive syndromes MKS and JEBTs in humans. In addition, there were cilia dependent Hedgehog signaling genes: DYNC2H1, GL12, GL13, GPR161, HHIP, IFT172, IFT57, IFT88, KIF7, MKS1, PRKACA, PTC1H, RPGRIP1L, SMO, and TULP3 present in MMC exomes. This study also found digenic heterogeneous URDVs in two cilia genes in 13 MMC subjects. Several subjects had two URDV-containing genes classified to sub-functional groups such as Smoothed signal regulation (i.e. C2CD3‑KIF7), Hedgehog off state (i.e. DYNC2H1‑RPGRIP1L), Hedgehog signaling (i.e. HHIP‑DYNC2H1), and dynein intermediate chain binding (i.e. DYNC2H1‑HTT). The extent of non-canonical WNT/PCP gene variants identified from subjects affected by various types of NTDs was recently published.19 This study identified 7% of total MMC subjects with one or more URDVs affecting the WNT/PCP signaling pathway and 3–4% affecting core PCP pathway genes involved in neural tube closure. A handful of WNT signaling genes were shown to have higher mutational burden in the current study cohort.19,20 Two URDVs [i.e. VANGL1 p.(V239I) and LRP6 p.(Y544C)] in two Mexican American subjects in the study cohort had been shown to cause functional loss of the proteins.20,21 Discovery of heterozygous digenic deleterious variants in human NTD cases with spina bifida (i.e. PTK7/VANGL1) and anencephaly (i.e. CELSR1/Scrib, CELSR1/DVL3) strongly support the strategy of screening PCP genes to discover risk alleles contributing to human NTDs.12,13 Here, we identified seven new digenic combinations of WNT/PCP signaling genes with three of these combinations (i.e. CELSR1‑LRP6, SMURF2‑LRP6, and PTK7‑VANGL1) involving the core PCP pathway. Of these, three subjects had URDVs potentially interfering with regulation of planar polarity establishment (GO:00090175). These new gene combinations suggest previously unknown gene–gene interaction candidates for validation with animal studies in the future.

Knowledge of the extent of genetic risk to ciliogenesis pathway function components to human MMC is limited. A quarter of the study subjects have one or more URDVs in 65 NTD candidate genes coding for ciliogenesis components to make microtubule structures: actomyosin filaments forming cell projections such as for validation with animal studies in the future.

### Table 1. Distribution of URDVs and genes in five biological ontology groups among MMC subjects. Note. Count of URDV, or neural tube defect candidate gene, or subjects with myelomeningocele are shown with the percentage to the total count of each column in bracket. Some genes are classified to more than one category. Count of URDVs, or gene or individual are presented follow by the percentage of the count presented in bracket. Some genes can be assigned to more than one ontology categories. A unique subtotal count shows the subtotal number of the URDV, or gene, or subject without double counting when a gene has multiple ontology groups assigned. Not in above – represent URDVs, genes or subjects outside the five ontologies. URDV—ultra-rare deleterious variant, EA—European American subject, MA—Mexican American subject. ECM – extracellular matrix.

| Categories | EA #URDV | #gene | #URDV/gene | #objects | MA #URDV | #gene | #URDV/gene | D#subjects |
|------------|---------|-------|------------|----------|---------|-------|------------|-----------|
| Cilium     | 63 (20.3%) | 35 (17.2%) | 1.80 | 54 (21.3%) | 53 (16.1%) | 33 (16.3%) | 1.61 | 49 (19.4%) |
| Cytoskeleton | 72 (23.2%) | 42 (20.6%) | 1.71 | 61 (24.0%) | 76 (23.0%) | 48 (23.6%) | 1.58 | 65 (25.8%) |
| ECM        | 46 (14.8%) | 22 (10.8%) | 2.09 | 40 (15.7%) | 38 (11.5%) | 16 (7.9%) | 2.38 | 35 (13.9%) |
| WNT-signaling | 44 (14.1%) | 31 (15.2%) | 1.42 | 41 (16.1%) | 51 (15.5%) | 32 (15.8%) | 1.59 | 46 (18.3%) |
| Cell migration | 77 (24.8%) | 52 (25.5%) | 1.48 | 70 (27.6%) | 66 (20.0%) | 44 (21.7%) | 1.50 | 55 (21.8%) |
| Unique subtotal | 185 (59.4%) | 113 (55.4%) | 1.63 | 128 (50.4%) | 181 (54.8%) | 110 (54.2%) | 1.65 | 125 (49.6%) |
| Other ontologies | 125 (40.6%) | 89 (44.1%) | 1.36 | 48 (18.9%) | 150 (45.2%) | 93 (45.8%) | 1.61 | 60 (23.8%) |
| No URDVs   | –       | –     | –          | 78 (30.7%) | –       | –     | –          | 67 (26.6%) |
| Total      | 310     | 202   | 1.56       | 254      | 331     | 203   | 1.63       | 252      |
as filopodia, lamellipodia, axoneme and cilia. Myosin motors along actin filaments facilitate vesicular cargo and biomolecule transport throughout the cell in response to the cues of signaling molecules transmitted from cilia\(^37\). In a study of 43 NTD trios, Lemay et al. (2015)\(^43\) identified heterozygote de novo deleterious variants in the \(\text{SHROOM3}\) gene of one MMC case and one anencephaly case. Only one MMC subject in this study had a \(\text{SHROOM3}\) URDV. Interestingly, half of the 65 cytoskeleton associated genes in MA and two-thirds in EA are associated with cytoskeleton and cilium, and the remaining genes belonging to actin cytoskeleton only. Digenic heterozygous URDV combinations found in MMC subject included \(\text{C2CD3/KIF7}\), \(\text{CC2D2A/WD11}\), \(\text{DYNC2H1/RPGRIP1L}\) and \(\text{HHT/DYNC2H1}\) affecting cytoskeleton and cilium; and \(\text{ABL1/CGN}\), \(\text{KIF20B/KATNAL2}\) and \(\text{MYH10/CGN}\) affecting actin cytoskeleton function.

Nearly 15% of the subjects in this study had URDVs in NTD candidate genes coding for ECM components. The ECM compartment serves as a microenvironment retaining growth factors and signaling molecules (e.g. \(\text{SHH, BMP, WNT, EGFs and FGFs}\)) to recruit or influence neuroepithelial cells to determine shape, size, polarity and migration status for neural tube closing\(^28\). The roles of ECM components on regulating the position of the nucleus, cell shape, proliferation, migration, and morphogenesis of neural tube has been well-discussed\(^1\). Loss of function in some ECM components (e.g. \(\text{Frem2 and Hspg2}\)) caused NTDs in animal models\(^9\). This study is the first to reveal URDVs in \(\text{FREM2}\), which were present in around 3% of study subjects, and URDVs in \(\text{HSPG2}\), which were present in 2% of study subjects. Pathogenic variants in \(\text{FREM2}\) have been associated with Cryptophthalmos (\#123,570) and Fraser syndrome 2 (\#617,666) with skull abnormalities or encephaloceles\(^44\). Digenic heterozygotes in two to four ECM genes were present in eight subjects in this study.

URDVs in cell migration genes were present in over 20% of the study subjects with many of these genes also involved in cilia, cytoskeleton, ECM, and WNT signaling. Ventral-dorsal migration of cells into the neural fold and rapid cell proliferation contribute to bending of the dorsolateral spinal neural plate, facilitating neural tube closure\(^45\). The \(\text{ct}\) mouse mutant has increased hindgut cell proliferation due to abnormal \(\text{Grhl3}\) expression at the caudal region which surpasses NE and NNE migration and proliferation and prevents spinal neural tube closure\(^5,46\). Knockdown of the cell migration molecule \(\text{Ighl1}\) prevents closure of the neurapore\(^15\). Mice with ablation of \(\text{Rac1}\), a molecule modulating actin cytoskeleton remodeling and cell migration in NNE, developed open spina bifida, exencephaly or anencephaly\(^16\). In addition, impaired mesodermal cell migration due to diabetic pregnancy led to NTDs in mouse\(^49\). Seventeen MMC subjects in this study had oligogenic heterozygous URDVs in genes involving in cell migration providing examples for testing oligogenic heterozygote effects on cell migration as a mechanism of MMC development.

Knowledge of oligogenic gene–gene interaction between genes of different ontologies contributing to risk of myelomeningocele is lacking. Oligogenic inheritance in holoprosencephaly, a rare birth defect of the brain, was recently demonstrated\(^18\). Evidence of mouse embryos with oligogenic heterozygous PCP related genes knockout developed spina bifida suggesting a subject having heterozygous URDVs in two NTD candidate genes

Figure 5. URDV containing candidate genes involved in multiple ontology groups. The Venn diagrams showed the relation and distribution of UTDV count and NTD candidate genes count to five ontology groups for EA (a) and MA (b). The number of URDVs and number of genes classified to cilium, WNT signaling, cytoskeleton, extracellular matrix (ECM) and cell migration are shown as fraction (count of URDV/count of gene). Genes classified to multiple categories were shown in the overlapped subset compartments between categories. Details on genes within each segment can be found in Supplementary Table 5.
| Subject  | Ethnicity | Gene Name | Variant ID (hg19) | aa Change | C-score | gnomAD exome controls | WNT signaling | cili um | EC M | cytoskeleton | cell migration | Relation to variant in same gene |
|----------|-----------|-----------|------------------|-----------|---------|----------------------|--------------|---------|-----|-------------|----------------|-------------------------------|
| F13_772  | EA        | NPHP3     | 3:1324008:34:A    | R130 5C   | 21.8    | √                   |              |         |     |             |                 |                                |
| F13_772  | EA        | HTT       | 4:3156038:A       | T117 3A   | 22       | 0                   |               | √       |     |             |                 |                                |
| F13_772  | EA        | FIGN      | 2:1644666:48:T    | K565 R    | 25.3    | 0                   |               |         |     |             |                 |                                |
| F13_772  | EA        | FREM2     | 13:392662:07:C    | P157 6A   | 24.3    | 0                   |               | √       |     | in cis     |                 |                                |
| F13_772  | EA        | FREM2     | 13:392662:11:T    | I157 7T   | 23.5    | 0                   |               | √       |     | in cis     |                 |                                |
| F13_772  | EA        | ALX1      | 12:856949:33:A    | I221 V    | 23.4    | 0                   |               |         |     |             |                 |                                |
| F36_1063 | EA        | FREM2     | 13:394207:83:C    | W20 31C   | 26.9    | 0                   |               |         |     |             |                 |                                |
| F36_1063 | EA        | HSPG2     | 1:2218211:4:C     | R191 9H   | 22.3    | 0                   |               |         |     |             |                 |                                |
| F36_1063 | EA        | FURIN     | 15:914215:18:T    | F275 S    | 23.2    | 0                   |               |         |     |             |                 |                                |
| F36_1063 | EA        | APOB      | 2:2125535:5:A     | I408 T    | 22.9    | 0                   |               | √       |     |             |                 |                                |
| F36_1063 | EA        | CC2D2     | 4:1557203:4:G     | R117 0K   | 24.1    | 0                   |               | √       |     |             |                 |                                |
| E18_859  | EA        | FLRT3     | 20:143070:42:G    | P371 S    | 22.1    | 0                   |               |         |     |             |                 |                                |
| E18_859  | EA        | NF1       | 17:296530:37:A    | I167 9V   | 21.1    | 0                   |               | √       |     |             |                 |                                |
| E18_859  | EA        | TIAM1     | 21:325823:99:C    | G784 R    | 24.4    | 0                   |               | √       |     |             |                 |                                |
| E18_859  | EA        | WDR11     | 10:122664:158:A   | T101 0A   | 24.5    | 0                   |               |         |     |             |                 |                                |
| E76_499  | EA        | RPGRI P1L | 16:536914:47:G    | A500 E    | 27.5    | 0                   |               | √       |     | in cis     |                 |                                |
| E76_499  | EA        | RPGRI P1L | 16:536914:41:T    | H502 L    | 27.7    | 0                   |               | √       |     | in cis     |                 |                                |
| E76_499  | EA        | MYH10     | 17:840269:5:C     | E128 2Q   | 26.6    | 0                   |               | √       |     |             |                 |                                |

Table 2. (continued)
| Subj ect | Ethnici ty | Genename | Variant ID (hg19) | aa change | gnomAD exomes | C-score | WNT signaling | ECM | Cytoskeleton | Cell migration | relation to variant in same gene |
|---------|------------|----------|------------------|------------|---------------|---------|---------------|-----|--------------|----------------|----------------------------------|
| E76_499 | EA         | MEGF8    | 19:42857105:G:A   | V1126M     | 22            | √       |               |     |              |                | √                               |
| A76_248 | EA         | TSC2     | 16:2134361:C:A     | Q1380K     | 7             | 0       | √             |     |              |                | √                               |
| A76_248 | EA         | CELSR1   | 22:46761272:C:T    | D2804N     | 24            | 0       | √             |     |              |                | √                               |
| A76_248 | EA         | APOB     | 2:21235259:G:A     | S1494L     | 26            | 0       | √             |     |              |                | √                               |
| D76_516 | EA         | SMO      | 7:128846046:A:G    | M326V      | 22            | 6       | √             |     |              |                | √                               |
| D76_516 | EA         | HIPK1    | 1:114515718:TA     | F1073I     | 24            | 4       | √             |     |              |                | √                               |
| D76_516 | EA         | PLOD3    | 7:100856185:A:C    | W273G      | 25            | 4       |               |     |              |                | √                               |
| C35_874 | EA         | C2CD3    | 11:73724483:C:T    | E2337K     | 24            | 5       | √             |     |              |                | √                               |
| C35_874 | EA         | KIF7     | 15:90174826:C:T    | R1004H     | 29            | 9       | √             |     |              |                | √                               |
| C35_874 | EA         | KIF7     | 15:90174827:C:A    | R1004C     | 33            | 0       | √             |     |              |                | √                               |
| C37_984 | EA         | FZD3     | 8:28384990:T:C     | I238T      | 25            | 1       | √             |     |              |                | √                               |
| C37_984 | EA         | HHIP     | 4:145629364:T:G    | M401R      | 20            | 9       |               |     |              |                | √                               |
| C37_984 | EA         | DYNC2H1  | 11:103107199:C:T   | Q3250H     | 22            | 6       | √             |     |              |                | √                               |
| E10_778 | EA         | SMURF2   | 17:62542045:G:A    | P723L      | 34            | 0       |               |     |              |                | √                               |
| E10_778 | EA         | LRP6     | 12:12277532:G:A    | S1505F     | 32            | 0       | √             |     |              |                | √                               |
| E10_778 | EA         | PTPR5    | 19:52194064:C:T    | E1280K     | 24            | 0       |               |     |              |                | √                               |
| E26_1036| EA         | RERE     | 1:84228752:C:A     | K590N      | 24            | 7       |               |     |              |                | √                               |
| E26_1036| EA         | KIF20B   | 10:91477207:T:C    | L360S      | 28            | 6       |               |     |              |                | √                               |
| E26_1036| EA         | KATNAL2  | 18:44627365:G:A    | E464K      | 27            | 0       |               |     |              |                | √                               |
| B81_701 | EA         | CELSR1   | 22:46859651:C:A    | R1379L     | 21            | 8       |               |     |              |                | √                               |
| B81_701 | EA         | LRP6     | 12:12317307:T:C    | N6515S     | 22            | 7       |               |     |              |                | √                               |
| F05_395 | EA         | HES1     | 3:193855988:C:T    | T270M      | 23            | 6       |               |     |              |                | √                               |

Table 2. (continued)
| subjec t | Ethnicity | genename | Variant ID (hg19) | aa change | C-score | gnamAD exoneme AF | cili um | WNT signaling | EC M | cytoskeleton | cell migration | relation to variant in same gene |
|----------|-----------|-----------|------------------|-----------|----------|------------------|--------|---------------|------|---------------|----------------|--------------------------------|
| F05_395  | EA        | NPHP3     | 3:1324051:07:A:C  | L110 9R   | 26.1     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| B73_616  | EA        | C2CD3     | 11:738050:05:C:T  | C106 7Y   | 26.2     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| B73_616  | EA        | SPECC1L   | 22:247248:17:C:T  | T717 1T   | 25.4     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| E16_856a | EA        | PDGFRA    | 4:5514355:7:C:T   | R357 1W   | 34       | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| E16_856a | EA        | GPR161    | 1:1680664:15:C:T  | A164 1T   | 22.9     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| E17_856b | EA        | PDGFRA    | 4:5514355:7:C:T   | R357 1W   | 34       | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| E17_856b | EA        | GPR161    | 1:1680664:15:C:T  | A164 1T   | 22.9     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| F23_948  | EA        | ABL1      | 9:1335898:39:C:T  | H45Y      | 26       | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| F23_948  | EA        | CGN       | 1:1514912:42:A:T  | K83X      | 36.0     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| E19_867  | EA        | LAMC1     | 1:1830952:84:A:G  | N944 1S   | 21.6     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| E19_867  | EA        | F2R       | 5:7602887:4:C:T   | S275 1F   | 23.4     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| A03_169  | EA        | ITGA6     | 2:1733523:24:G:A  | C737 1Y   | 31       | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| A03_169  | EA        | APOB      | 2:2123263:4:T:A   | K236 9M   | 23.5     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| A55_1079 | EA        | IFT172    | 2:2767698:5:T:C   | E119 2G   | 27       | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| A55_1079 | EA        | DVL1      | 1:1275803:24:C:G  | R229 1P   | 20.6     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| B63_581  | EA        | HTT       | 4:3180083:24:C:T  | R150 8C   | 28.1     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| B63_581  | EA        | DYNC2H1   | 11:102988:518:T:C | Y309 1H   | 22.6     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| B87_728  | EA        | PTC1      | 9:9823209:6:T:C   | S616 1G   | 23.4     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| B87_728  | EA        | KIAA0586  | 14:589252:64:G:A  | c.181 5+1 | 24.7     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| C19_810  | EA        | COBL      | 7:5108525:3:C:A   | V127 1L   | 22.5     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| C19_810  | EA        | RAPGRI1   | 16:536988:95:T:G  | Q377 1P   | 29.6     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| D81_607  | EA        | GLI3      | 7:4205337:0:G:A   | N110 1D   | 24       | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |
| D81_607  | EA        | ALX1      | 12:856773:89:A:G  | H89R      | 22.6     | ✓                | ✓      | ✓             | ✓    | ✓             | ✓              | ✓                |

Table 2. (continued)
Table 2. (continued)

| Subject | Ethnicity | Gene Name | Variant ID (hg19) | aa change | C-score | gnomAD exome %AF | ci | WNT signaling | ECM | cytoskeleton | cell migration | relation to variant in same gene |
|---------|-----------|-----------|------------------|-----------|---------|-----------------|---|----------------|-----|--------------|--------------|-------------------------------|
| E87_244 | EA        | FREM2     | 13:394335         | G244S     | 32      | 0               | √ |                |     |              |              |                               |
|         |           |           | 41:G:A           |           |         |                 |   |                |     |              |              |                               |
| E87_244 | EA        | FLRT3     | 20:143068         | P447L     | 23.8    | 0               | √ |                |     |              |              |                               |
|         |           |           | 13:G:A           |           |         |                 |   |                |     |              |              |                               |
| F21_872 | EA        | DYN2H1    | 11:103124         | T340A     | 33      |                | √ |                |     |              |              |                               |
|         |           |           | 154:A:G          |           |         |                 |   |                |     |              |              |                               |
| F21_872 | EA        | RPRG1P1L  | 16:536360         | G128V     | 24.4    | 0               | √ |                |     |              |              |                               |
|         |           |           | 79:C:A           |           |         |                 |   |                |     |              |              |                               |
| F63_305 | MA        | AGO2      | 8:1415953         | G35R      | 24.7    | 0               | √ |                |     |              |              |                               |
|         |           |           | 30:C:T           |           |         |                 |   |                |     |              |              |                               |
| F63_305 | MA        | SALL1     | 16:511760         | K40N      | 22.4    | 0               | √ |                |     |              |              |                               |
|         |           |           | 13:C:A           |           |         |                 |   |                |     |              |              |                               |
| F63_305 | MA        | IFT17      | 2:2767069         | L151      | 22.7    | 0               | √ |                |     |              |              |                               |
|         |           |           | 0:G:A            |           |         |                 |   |                |     |              |              |                               |
| F63_305 | MA        | FREM2      | 13:392660         | T152      | 27.7    | 0               | √ |                |     |              |              |                               |
|         |           |           | 42:A:C           |           |         |                 |   |                |     |              |              |                               |
| D53_759 | MA        | VANGL1     | 1:1162279         | Q376R     | 22.8    | 0               | √ |                |     |              |              |                               |
|         |           |           | 61:A:G           |           |         |                 |   |                |     |              |              |                               |
| D53_759 | MA        | PTK7       | 6:4309976         | R284      | 23.6    | 0               | √ |                |     |              |              |                               |
|         |           |           | 8:G:A            |           |         |                 |   |                |     |              |              |                               |
| D53_759 | MA        | PTPRF      | 1:4405676         | G358      | 21.3    | 0               | √ |                |     |              |              |                               |
|         |           |           | 6:G:T            |           |         |                 |   |                |     |              |              |                               |
| D53_759 | MA        | AMBR1      | 11:465638         | N559      | 22.2    | 0               | √ |                |     |              |              |                               |
|         |           |           | 91:T:C           |           |         |                 |   |                |     |              |              |                               |
| E35_115 | MA        | KITLG      | 12:889262         | K56R      | 22.2    | 0               | √ |                |     |              |              |                               |
|         |           |           | 43:T:C           |           |         |                 |   |                |     |              |              |                               |
| E35_115 | MA        | CFL1       | 11:656236         | G10       | 23.3    | 0               | √ |                |     |              |              |                               |
|         |           |           | 88:CA            |           |         |                 |   |                |     |              |              |                               |
| E35_115 | MA        | TBX5       | 6:1665719         | E402      | 23.7    | 0               | √ |                |     |              |              |                               |
|         |           |           | 07:C:T           |           |         |                 |   |                |     |              |              |                               |
| E35_115 | MA        | TCF7L1     | 2:8553358         | N386      | 23.5    | 0               | √ |                |     |              |              |                               |
|         |           |           | 3:CA             |           |         |                 |   |                |     |              |              |                               |
| F61_292 | MA        | NF1        | 17:295597         | R113      | 24.5    | 0               | √ |                |     |              |              |                               |
|         |           |           | 97:C:T           |           |         |                 |   |                |     |              |              |                               |
| F61_292 | MA        | MYH10      | 17:838753         | S169      | 22.9    | 0               | √ |                |     |              |              |                               |
|         |           |           | 9:A:T            |           |         |                 |   |                |     |              |              |                               |
| F61_292 | MA        | CGN        | 1:1515025         | G755      | 22.2    | 0               | √ |                |     |              |              |                               |
|         |           |           | 42:A:G           |           |         |                 |   |                |     |              |              |                               |
| F61_292 | MA        | BRCA2      | 13:329688         | V309      | 21.2    | 0               | √ |                |     |              |              |                               |
|         |           |           | 40:A:G           |           |         |                 |   |                |     |              |              |                               |
| F80_426 | MA        | CC2D2      | 4:1559123         | H141      | 21.8    | 0               | √ |                |     |              |              |                               |
|         |           |           | 4:C:A            |           |         |                 |   |                |     |              |              |                               |
| F80_426 | MA        | WDR11      | 10:122646         | N672      | 24.4    | 0               | √ |                |     |              |              |                               |
|         |           |           | 241:A:T          |           |         |                 |   |                |     |              |              |                               |
| F80_426 | MA        | HSPG2      | 1:2221449         | V213      | 22.2    | 0               | √ |                |     |              |              |                               |
|         |           |           | 7:C:T            |           |         |                 |   |                |     |              |              |                               |

Scientific Reports | (2021) 11:3639 | https://doi.org/10.1038/s41598-021-83058-7 | natureportfolio
| Subject | Ethnicity | Gene Name | Variant ID (hg19) | Aa change | C-score | gnomAD Exome Center AF | WNT signaling | ECM | Cytoskeleton | Cell Migration |
|---------|-----------|-----------|------------------|-----------|---------|------------------------|---------------|-----|--------------|---------------|
| A14_285 | MA        | DLC1      | 8:1294613:8:G:A  | R1384C    | 34      | 0                      | ✓             | ✓   | ✓            |               |
| A14_285 | MA        | GBX2      | 2:2370746:49:G:C  | R319G     | 22.3    | ✓                      |               | ✓   | ✓            |               |
| A14_285 | MA        | FREM2     | 13:392632:88:C:T  | P603S     | 20.4    | ✓                      |               | ✓   | ✓            |               |
| D25_473 | MA        | HHIP      | 4:1456558:94:C:T  | P588S     | 21.8    | 0                      | ✓             | ✓   | ✓            |               |
| D25_473 | MA        | SCUB1     | 22:436187:13:G:T  | P455T     | 22.8    | ✓                      | ✓             | ✓   | ✓            |               |
| D25_473 | MA        | SIRT1     | 10:696760:75:G:A  | V657I     | 23.2    | ✓                      | ✓             | ✓   | ✓            |               |
| E36_132 | MA        | MKKS      | 20:103939:36:G:A  | I76T      | 23.5    | ✓                      | ✓             | ✓   | ✓            |               |
| E36_132 | MA        | DLC1      | 8:1325905:9:G:C   | L365V     | 26.2    | ✓                      | ✓             | ✓   | ✓            |               |
| E36_132 | MA        | APOB      | 2:2123031:4:G:C   | I3142M    | 21.5    | ✓                      | ✓             | ✓   | ✓            |               |
| D09_398 | MA        | STK4      | 20:436237:91:G:A  | I196V     | 25.5    | ✓                      | ✓             | ✓   | ✓            |               |
| D09_398 | MA        | NPHELL3   | 3:1324008:40:G:A  | P13035    | 21.8    | 0                      | ✓             | ✓   | ✓            |               |
| D09_398 | MA        | MYO10     | 5:1670166:2:C:T   | D948N     | 22.5    | 0                      | ✓             | ✓   | ✓            |               |
| D28_502 | MA        | TME6Z     | 8:9480957:4:G:A   | R659Q     | 22.5    | 0                      | ✓             | ✓   | ✓            |               |
| D28_502 | MA        | VCL       | 10:758746:25:G:T  | R1076W    | 29.5    | 0                      | ✓             | ✓   | ✓            |               |
| D28_502 | MA        | NOTCH1    | 9:1394131:3:G:C   | G347R     | 25.4    | ✓                      | ✓             | ✓   | ✓            |               |
| D36_600 | MA        | PKN2      | 1:8923711:7:G:A   | I212M     | 23.6    | ✓                      | ✓             | ✓   | ✓            |               |
| D36_600 | MA        | DLC1      | 8:1295708:8:C:T   | V920I     | 23.2    | ✓                      | ✓             | ✓   | ✓            |               |
| D36_600 | MA        | PRKA CA   | 19:142040:36:G:A  | F315Y     | 22.7    | ✓                      | ✓             | ✓   | ✓            |               |
| A83_383 | MA        | CORO1C    | 12:109051:112:T:C | M293V     | 22.9    | 0                      | ✓             | ✓   | ✓            |               |
| A83_383 | MA        | ABI1      | 10:270592:19:G:A  | T178I     | 24.9    | ✓                      | ✓             | ✓   | ✓            |               |
| A19_424 | MA        | FREM2     | 13:394241:82:G:C  | K2129N    | 23.3    | 0                      | ✓             | ✓   | ✓            |               |
| A19_424 | MA        | PTPRF     | 1:4405755:7:C:G   | P536A     | 21.4    | 0                      | ✓             | ✓   | ✓            |               |
| A27_495 | MA        | DISP1     | 1:2231763:79:G:A  | R547H     | 22.7    | 0                      | ✓             | ✓   | ✓            |               |

Table 2. (continued)
belonging to the same ontology group may increase risk of MMC development. So far, no non-syndromic human spina bifida cases with homozygous deleterious variants in one gene has been found, and only cases with digenic heterozygous rare deleterious variants in two PCP genes had been reported. This study cohort showed examples of MMC subjects carrying multiple NTD candidate genes containing URDVs and one subject has two compound heterozygous URDVs in the \( KIF7 \) gene. Figure 6 summarized the potential impact of URDVs in NTD candidate genes identified from 506 MMC subjects. Hypothesis-free testing of random combinations of gene–gene interaction in animal models is too labor intensive and cost-prohibitive. While MMC risk for some subjects might be rationalized by disrupted genetic interaction between URDV-containing genes within the same ontology, subjects with URDV disrupted genes in different ontologies will be challenging. Here, the study results provide examples of new highly probable genetic interacting partners elevating MMC risk through an oligogenic

| subject | Ethnicity | gene name | variant ID (hg19) | aa change | C-score | gnomAD exome AF | clinum | WNT signaling | ECM | cytoskeleton | cell migration | relation to variant in same gene |
|---------|-----------|-----------|------------------|------------|---------|----------------|--------|---------------|-----|-------------|----------------|----------------------|
| A27_495 | MA        | HTT       | 4:3189379;C:T    | A166_4V    | 23      | 0             | √      |               |     |             |                |                      |
| E52_879 | MA        | AGO2      | 8:1415663;40:C:T | D358_N     | 24      | 0             |         |               |     |             |                |                      |
| E52_879 | MA        | LRP6      | 12:123181;44:C:T | Y544_C     | 27      | 0             |         |               |     |             |                |                      |
| D38_634 | MA        | TBC1      | 6:1215444;34:T:A | D810_V     | 21      | 0             |         |               |     |             |                |                      |
| D38_634 | MA        | DYN2H1    | 11:103033;832:C:T| R152_3G    | 22      | 0             |         |               |     |             |                |                      |
| D51_724 | MA        | KIF20B    | 10:914749;01:G:T | R301_L     | 29      | 0             |         |               |     |             |                |                      |
| D51_724 | MA        | STIL      | 1:4774619;2:A:T  | S646_R     | 20      | 0             |         |               |     |             |                |                      |
| D59_862 | MA        | DLC1      | 8:1335691;9:G:T  | P221_H     | 23      | 0             |         |               |     |             |                |                      |
| D59_862 | MA        | SEMA4C    | 2:9753351;7:C:A  | G36_V      | 22      | 0             |         |               |     |             |                |                      |
| D05_381 | MA        | HSPG2     | 1:2216600;7:C:T  | R324_9H    | 27      | 0             |         |               |     |             |                |                      |
| D05_381 | MA        | ITGA3     | 17:481537;98:C:A | P595_T     | 24      | 0             |         |               |     |             |                |                      |
| F56_260 | MA        | MDM2      | 12:692332;29:T:C | V365_A     | 28      | 0             |         |               |     |             |                |                      |
| F56_260 | MA        | PDGF C    | 4:1576843;01:CA  | V327_L     | 22      | 0             |         |               |     |             |                |                      |
| F76_355 | MA        | CALR      | 19:130543;88:CA  | T333_N     | 27      | 0             |         |               |     |             |                |                      |
| F76_355 | MA        | APOB      | 2:2123275;1:A:T  | V233_0E    | 28      | 0             |         |               |     |             |                |                      |

Table 2. Subjects with multiple URDVs in candidate genes classified with one or more ontological groups. Note: EA—European American subjects, MA—Mexican American subjects. URDV—ultra-rare deleterious single nucleotide variant defined in Materials and Method section. Gene name—is the symbol of gene recommended by the Human Genome Organization Gene Nomenclature Committee (HGNC). Gene assigned into cilium, WNT-signaling, extracellular matrix (ECM), cytoskeleton and cell migration was described in Materials and Methods, and Supplementary Table 3. C-Score—Combined Annotation Dependent Depletion (CADD) score is a broadly applicable metric that objectively weights and integrates diverse information that integrates multiple functional annotations into one metric by contrasting variants that survived natural selection with simulated mutations.
heterozygote mechanism. For example, subject E52-879 has URDVs p.(D358N) in AGO2 and p.(Y544C) in LRP6. While interaction of AGO2 and LRP6 has not yet been demonstrated, it has been shown that inactivated canonical WNT signaling facilitated nuclear entry of miR-133a and complexed with AGO2 to suppress DnmT3b expression in mouse HL-1 cells. Also, the presence of p.Y544C impaired LRP6 localization to plasma membrane for facilitating WNT signaling. Subject D22-469 carries URDVs p.(N35K) and p.(H31R) respectively in SNX3 and CITED2, known components for WNT signaling and cell migration respectively. SNX3 binds WLS to regulate recycling of WLS thereby regulating the level of WNT excretion. Furthermore, it has been demonstrated that SNX3 with the subject’s URDV p.(N35K) failed to interact with WLS affecting WNT sorting and secretion. A recent report demonstrated the developmental defects of Cited2 morphants in Zebrafish could be rescued with exogenous WNT5A and WNT11.

This study identified URDVs in NTD candidate genes in 70% of MMC subjects lending strong support to the approach of screening NTD candidate genes to identify high impact genetic risk factors associated with human MMC development. Results of oligogenic URDV-containing cases present in this study provide a rational basis to test oligo-genic combinations in animal models to elucidate the mechanism(s) of MMC in humans. New oligogenic disease mechanism(s) for MMC development can be discovered from the new combinations of genes presented in this study and not previously known to cause NTDs. Testing genetic interactions between URDV-containing NTD candidate genes discovered from MMC subjects to understand human MMC mechanism is sound and promising.

In the study, we observed the presence of URDV-containing candidate genes that were common to EA and MA, and unique to one or the other ethnic population. Many of the genes unique to EA or MA belongs to the same ontology subclasses and a small number of genes belong to distinct specific subclasses. Differences in subclass information may help to reveal the disease mechanisms for the subjects who carried these variants. The potential of utilizing the observed differences in these genetic variants to predict differential risks in different ethnic populations may require replication with additional studies.

The study results were limited to examining the impact of ultra-rare SNVs in the coding region and the splice sites of NTD candidate genes. The impact of URDVs in genes not known to cause NTDs in mice has not been evaluated. Yet there are a significant portion of genes which have important biological function and are detected in the developmental stages when the neural tube closes that can affect neural tube development. Importance of rare deleterious variants, out of frame indel variants, deleterious non-coding variants and copy number variants should be recognized. Previous genetic association studies demonstrated some impact of common and rare deleterious variants to NTD risk in humans but generally lack replication due to various limitations of association study designs, ethnic background, and sample size. This study cohort has similar limitations. Separate analysis using all the common variants identified from the subject exomes did not yield significant association with the selected NTD candidate genes after correcting for multiple testing.
Materials and methods

Study population. A total of 511 subjects were selected for whole exome sequencing (WES) from a myelomeningocele study cohort enrolled from spina bifida clinics in five locations of North America between 1997 and 2010 (Au et al., 2008)39. All subjects provided an informed consent and enrolled in accordance with an institutional IRB at the University of Texas Health Science Center (UTHealth) at Houston. WES was performed on DNAs of 257 EA MMC including 140 females and 117 males, and DNAs of 254 MA MMC including 134 females and 120 males. Subjects in the study are sporadic cases and reported no family history of MMC. The majority (81%) of the study subjects were born before January 1998. The date North American countries mandated fortification of food crops with folic acid43. Fifty-two subjects were born in 1998 and 46 were born after 1998. Approximately 24% of the study subjects have MMC lesions at or above vertebrae L1 up to T10, 66.7% at or below vertebrae L2 down to sacrum, 1.7% had lesion spanning L4 to T10, and 8.2% had no specific lesion level information. Around 60% of the study subjects had hydrocephaly, 5.5% had arrested hydrocephaly, 3.7% had no hydrocephaly and 31.1% had no specific information. The majority subjects (74.8%) did not have information on Chiari Malformation, 0.6% had type I and 22.5% had type I Chiari Malformation, and 2.2% were normal. Blood samples were collected from subjects and parents where possible and genomic DNA was extracted for the study.

Exome sequencing and variant annotation. Exome library probes were made from an in-house design based on TargetSeq (Life Technologies, Inc.) with addition of splice sites, UTRs, small non-coding RNAs (e.g., microRNAs) and a selection of miRNA binding sites, and 200 bp promoter regions. High quality genomic DNA samples were processed using the exome library probes and the captured DNA products were sequenced following the manufacturer's standard protocol for multiplexed sequencing using the P1 chip on the Ion Proton platform (Life Technologies, Inc.). Quality of sequencing was maintained at 40–60 million reads with read length between 120 and 150 bases, and over 75% reads were on-target for all successfully sequenced samples. Other quality control measures were implemented to map 45–60,000 SNP per sample with ~ 50% heterozygote variants and the Ts/Tv ratio average of 2.5. Samples that failed to meet the above quality criteria were re-sequenced. If re-sequencing failed, another subject’s DNA will be used to reach the goal of sequencing at least 500 subjects.

For sequence data that passed variant level and sample level quality filters, variant calling was conducted using Genome Analysis Toolkit GATK HaplotypeCaller version 3.x following best practice guidelines. Additional variant filtering steps had been described previously39,40. Briefly, only variants designated a “PASS” by Variant Quality Score Recalibration (VQSR), met map quality score < 20, and having inbreeding coefficient < -0.3 were retained for further analysis. Individual sample filters were used to ensure that only high-fidelity variants with alternate allele depth > 25%, a read depth > 10, and genotype quality score > 20 were analyzed. Allele count (AC), allele number (AN), and allele frequency (AF) were recalculated for individual ethnicities after the filtering processes. Filtered high quality single-nucleotide variants (SNVs) were annotated using the non-synonymous single-nucleotide variant functional predictions (dbNSFP)76 version 4.0 database for all functional prediction information publicly available. Further analysis was focused on SNVs leading to stop_gained, stop_lost, non-synonymous, splice donor, and splice acceptor site changes in the canonical transcripts.

Thirty-three EA subjects and 55 MA subjects in the study were also re-sequenced using Illumina NGS platform (i.e. NovaSeq). Raw sequence reads of WES were mapped to GRCh38 (hg38) and read for WGS were mapped to GRCh37 (hg19). Variants were called following the GATK best practice guidelines. Then the variant data generated was filtered using the same filtering criteria described above for the Ion Proton generated data. Filtered variant data of the 88 subjects generated by the Ion Proton platform used in the study were extracted and compared to filtered variant data generated by the Illumina Platform to verify variants with concordance.

Ultra-rare functional deleterious SNVs (URDVs) analysis. The alternate allele frequencies (aAFs) of variants reported for the exomes of the non-Finnish European (NFE) and Ad Mixed American (AMR) control populations in the genome aggregation database gnomAD v2.1.1 (https://gnomad.broadinstitute.org/) were used to annotate the single nucleotide variants (SNVs) in the study33. Variants having ethnic aAF == 0 or absent in gnomAD v2.1.1 (controls) despite high location coverage in gnomAD were defined as ultra-rare SNVs (URVs). Variants having aAFs in NFE or AMR gnomAD v2.1.1 (controls) less than 0.01 were defined as rare, whereas aAF ≥ 0.01 were defined as common. Combined Annotation Dependent Depletion (CADD; https://cadd.gs.washington.edu/) phred scores (C-score) of variants were used as the model to predict deleteriousness and variants with a C-score ≥ 20 (top 1% most deleterious) cutoff were defined as ultra-rare deleterious variants (URDVs) and retained for further comparison in the study. Genes with URDVs were further evaluated for potential association with MMC in the study cohort using the in silico function analysis tools. In addition, all URDVs identified in the NTD candidate genes in Supplementary Table 3 had been manually examined and confirmed using IGV Web App (https://igv.org/app/)34.

Online variant analysis tools. Annotation of gene and function with Gene Ontology59 (GO; http://geneontology.org) terms were used to estimate the extent of enrichment for particular ontology in molecular function, biological process, cell components and pathways in each group of genes found with URDVs to evaluate their potential relevance to development of neural tube defects. Several web based tools were used including PubMed and Mouse Genome Database5 (http://www.informatics.jax.org/phenotypes.shtml) to compile a list of genes known to associate with neural tube defects (Supplementary Table 2) searching for terms: neural tube defects, open neural tube, spina bifida, anencephaly, exencephaly, craniorachischisis, and abnormal neural tube morphology. ToppFun (https://toppgene.cchmc.org/enrichment.jsp) and ToppCluster50 (https://toppcluster.cchmc.org/) were used to evaluate ontology enrichment and for comparing multiple groups of genes (Chen et al. 2009). Genes in the GO terms were extracted using g:Profiler50 (https://biit.cs.ut.ee/gprofiler/convert). Gene tolerant
to loss-of-function mutation (pLI score) and haploinsufficiency were annotated using resources from Database of genomic Variation and Phenotype in Humans using Ensembl Resources (DECIPHER, https://decipher.sanger.ac.uk) [45].

To annotate expression of the genes of interest in animal neural tube, a database for human neural tubes during Carnegie stages CS12 and CS13 [46] (Krupp et al., 2008) was used. Genes expressed in mouse future spinal cord between Theliler Stages TS11-19 and early embryonic lethality mouse genes (MP:0008762) were extracted from MGI (http://www.informatics.jax.org/) for function annotation.

**NTD candidate genes.** Increasing numbers of published reports showed compound heterozygote deleterious variants of mouse NTD genes in human NTD subjects suggesting these variants as genetic risk factors for human NTDs. To facilitate variant containing gene prioritization, a list of 551 genes (Supplementary Table 2) was assembled from Mouse Genome Database [9] (http://www.informatics.jax.org/) using search terms: open neural tube (MP:00000929), exencephaly (MP:0009914), anencephaly (MP:0001890), craniorachischisis (MP:0008784), and abnormal neural tube closure (MP:0003720). Also included were RARA, RARG, KIAA0586 and KATNAL2 with evidence of open neural tube defects in mice, chicken, frog and human [62-64]. Mouse NTD mutants not assigned to a gene and genes assigned only to spina bifida occulta were not included in this study. In addition, a list of 14 genes were added because association has been established in our study cohort (Supplementary Table 2). Together, a total of 551 candidate genes were selected for the study.

**Ethical compliance.** Affected subjects and/or their parents were recruited for the research study with a consenting protocol approved by the University of Texas Health Science Center (UTHealth) at Houston Institutional Review Board.

**Data availability**

Variants information is present in Supplementary Table 3. Also, variants will be available for review upon publication at dbSNP (https://www.ncbi.nlm.nih.gov/snp/) under BioProject ID: PRJNA611755.

Received: 6 November 2020; Accepted: 28 January 2021 Published online: 11 February 2021

**References**

1. Nikolopoulou, E., Galea, G. L., Rolo, A., Greene, N. D. E. & Copp, A. J. Neural tube closure: Cellular, molecular and biomechanical mechanisms. Development 144, 552–566 (2017).

2. Veland, I. R., Lindbæk, L. & Christensen, S. T. Linking the primary cilium to cell migration in tissue repair and brain development. Bioessays 41, 1115–1125 (2014).

3. Juriloff, D. & Harris, M. Insights into the etiology of mammalian neural tube closure defects from developmental, genetic and evolutionary studies. J. Dev. Biol. 6, 22 (2018).

4. Ray, H. J. & Niswander, L. A. Grainyhead-like 2 downstream targets act to suppress epithelial-to-mesenchymal transition during neural tube closure. Development 143, 1192–1204 (2016).

5. Gustavsson, P. et al. Increased expression of Grainyhead-like-3 rescues spina bifida in a folate-resistant mouse model. Hum. Mol. Genet. 16, 2640–2646 (2007).

6. Harris, M. J. & Juriloff, D. M. An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. Birth Defects Res. Part A Clin. Mol. Teratol. 88, 601–611 (2010).

7. Salbaum, J. M. & Kappen, C. Neural tube defect genes and maternal diabetes during pregnancy. Birth Defects Res. Part A Clin. Mol. Teratol. 88, 601–611 (2010).

8. Eppig, J. T. et al. The Mouse Genome Database (MGD): Comprehensive resource for genetics and genomics of the laboratory mouse. Nucleic Acids Res. https://doi.org/10.1093/nar/gkr974 (2012).

9. Blake, J. A., Butch, C. J., Kadin, J. A., Richardson, J. E. & Eppig, J. T. The Mouse Genome Database (MGD): Premier model organism resource for mammalian genomics and genetics. Nucleic Acids Res. 39, D842–D848 (2011).

10. Chen, Z. et al. Genetic analysis of Wnt/PCP genes in neural tube defects. BMC Med. Genomics 11, 38 (2018).

11. Ishida, M. et al. A targeted sequencing panel identifies rare damaging variants in multiple genes in the cranial neural tube defect, anencephaly. Clin. Genet. 93, 870–879 (2018).

12. Wang, L. et al. Digenic variants of planar cell polarity genes in human neural tube defect patients. Mol. Genet. Metab. 124, 94–100 (2018).

13. Lemay, P. et al. Whole exome sequencing identifies novel predisposing genes in neural tube defects. Mol. Genet. Genomic Med. 7, e00467 (2019).

14. Bai, E. L., Cheng, A. N. & Sankaran, V. G. The genetics of human hematopoiesis and its disruption in disease. EMBO Mol. Med. 11, e10316 (2019).

15. Katsanis, N. et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science (80‑.) 363, D886–D894 (2019).

16. Harris, M. J. & Juriloff, D. M. Mouse mutants with neural tube closure defects and their role in understanding human neural tube defects. Birth Defects Res. Part A Clin. Mol. Teratol. 79, 187–210 (2007).

17. Kim, A. et al. Integrated clinical and omics approach to rare diseases: Novel genes and oligogenic inheritance in holoprosencephaly. Brain https://doi.org/10.1093/brain/awy290 (2019).

18. Wang, M., Capra, & Kibar. Update on the role of the non-canonical Wnt/planar cell polarity pathway in neural tube defects. Cells 8, 1198 (2019).

19. Kaimal, V., Bardes, E. E., Tabar, S. C., Jegga, A. G. & Aronow, B. J. ToppCluster: A multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. Nucleic Acids Res. 38, W96–W102 (2010).

20. Hillman, P. et al. Identification of novel candidate risk genes for myelomeningocele within the glucose homeostasis/oxidative stress and folate/one-carbon metabolism networks. Mol. Genet. Genomic Med. 8(11), e1495. https://doi.org/10.1002/mg3.1495 (2020).

21. Krupp, D. R. et al. Transcriptome profiling of genes involved in neural tube closure during human embryonic development using long serial analysis of gene expression (long-SAGE). Birth Defects Res. Part A Clin. Mol. Teratol. 94, 683–692 (2012).
23. Vogel, T. W., Carter, C. S., Abode-lymah, K., Zhang, Q. & Robinson, S. The role of primary cilia in the pathophysiology of neural tube defects. *Neuromus. Focus* 33, E2 (2012).
24. Christensen, S. T., Pedersen, S. F., Satir, P., Veland, I. R. & Schneider, L. Chapter 10 the primary cilium coordinates signaling pathways in cell cycle control and migration during development and tissue repair. In *Current Topics in Developmental Biology* 261–301 (2008). https://doi.org/10.1016/S0070-2153(08)00010-7.
25. Murdoch, J. N. & Copp, A. J. The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. *Birth Defects Res. Part A Clin. Mol. Teratol.* 88, 633–652 (2010).
26. Ybot-Gonzalez, P., Cogram, P., Gerrelli, D. & Copp, A. J. Sonic hedgehog and the molecular regulation of mouse neural tube closure. *Development* 129, 2507–2517 (2002).
27. Ybot-Gonzalez, P. et al. Convergent extension, planar–cell polarity signalling and initiation of mouse neural tube closure. *Development* 134, 789–799 (2007).
28. Anvarian, Z., Mykytyn, K., Mukhopadhyay, S., Pedersen, L. B. & Christensen, S. T. Cellular signalling by primary cilia in development, organ function and disease. *Nat. Rev. Neurosci.* 15, 199–219 (2019).
29. Wheway, G., Nazlamova, L. & Hancock, J. T. Signalling through the primary cilium. *Front. Cell Dev. Biol.* 6, 8. https://doi.org/10.3389/fcel.2018.00008 (2018).
30. Chen, X. et al. Detection of Copy Number Variants Reveals Association of Cilia Genes with Neural Tube Defects. *PLoS ONE* 8, e54492 (2013).
31. Rolo, A., Esquin, S., Greene, N. D. E. & Copp, A. J. Rho GTPases in mammalian spinal neural tube closure. *Small GTPases* 9, 283–289 (2018).
32. Malicki, J. J. & Johnson, C. A. The cilium: Cellular antenna and central processing unit. *Trends Cell Biol.* 27, 126–140 (2017).
33. Gurniak, C. B., Perlas, E. & Witke, W. The actin depolymerizing factor n-cofilin is essential for neural tube morphogenesis and neuronal crest cell migration. *Dev. Biol.* 278, 231–241 (2005).
34. Wallingford, J. B. Planar cell polarity, ciliogenesis and neural tube defects. *Hum. Mol. Genet.* 15, R227–R234 (2006).
35. May-Simera, H. L. & Kelley, M. W. Cilia, Wnt signaling, and the cytoskeleton. *Front. Cell Dev. Biol.* 3459–3469 (2010).
36. Mirvis, M., Stearns, T. & James Nelson, W. Cilium structure, assembly, and disassembly regulated by the cytoskeleton. *Biochem. J.* 475, 2329–2353 (2018).
37. Simons, M. et al. Inversion, the gene product mutated in nephropathohism type II, functions as a molecular switch between Wnt signaling pathways. *Nat. Genet.* 37, 537–543 (2005).
38. Beaumont, M. et al. Targeted panel sequencing establishes the implication of planar cell polarity pathway and involves new candidate genes in neural tube defect disorders. *Hum. Genet.* 138, 363–374 (2019).
39. Hebert, H. et al. Burden of rare deleterious variants in WNT signaling genes among 511 myelomeningocele patients. *PLoS One* 15(9), 83. https://doi.org/10.22541/au.15983195.75483136 (2020).
40. Reynolds, A. et al. VANG1 rare variants associated with neural tube defects affect convergent extension in zebrafish. *Mech. Dev.* 127, 385–392 (2010).
41. Lei, Y. et al. Rare LRPS6 Variants Identified in Spina Bifida Patients. *Hum. Mutat.* 36, 342–349 (2015).
42. Lemay, P. et al. Loss-of-function de novo mutations play an important role in severe human neural tube defects. *J. Med. Genet.* 52, 493–497 (2015).
43. van Haelst, M. M., Scambler, P. J. & Hennekam, R. C. M. Fraser syndrome: A clinical study of 59 cases and evaluation of diagnostic criteria. *Am. J. Med. Genet. Part A* 143A, 3194–3203 (2007).
44. McShane, S. G. et al. Cellular basis of neuroepithelial bending during mouse spinal neural tube closure. *Dev. Biol.* 404, 113–124 (2015).
45. Brook, F. A., Shum, A. S. W., Van Straaten, H. W. M. & Copp, A. J. Curvature of the caudal region is responsible for failure of neural tube closure in the curly tail (ct) mouse embryo. *Development* 113, 671–678 (1991).
46. Morita, H. et al. Cell movements of the deep layer of non-neural ectoderm underlie complete neural tube closure in Xenopus. *Development* 139, 1417–1426 (2012).
47. Rolo, A., Galea, G. L., Savery, D., Greene, N. D. E. & Copp, A. J. Novel mouse model of ependymoma: Post-neuropilization origin and relationship to open neural tube defects. *Dis. Model. Mech.* 12, dmm04083 (2019).
48. Salbaum, J. M. et al. Novel mode of defective neural tube closure in the non-obese diabetic (NOD) mouse strain. *Sci. Rep.* 5, 16917 (2015).
49. Di Mauro, V., Crasto, S., Colombo, E. S., Di Pasquale, E. & Catalucci, D. Wnt signalling mediates miR-133a nuclear re-localization for the transcriptional control of Dmnt3b in cardiac cells. *Sci. Rep.* 9, 9320 (2019).
50. Brown, H. M., Murray, S., Northrup, H., Au, K. S. & Niswander, L. A. Snx3 is important for mammalian neural tube closure via its role in canonical and non-canonical WNT signaling. *Development* 147(22), dev192518 (2020).
51. Santos, J. M. A. et al. Exogenous WNT3A and WNT11 proteins rescue CITED2 dysfunction in mouse embryonic stem cells and zebrafish morphants. *Cell Death Dis.* https://doi.org/10.1038/s41419-019-1816-6 (2019).
52. Fu, K. S. et al. Characteristics of a spina bifida population including North American Caucasian and Hispanic individuals. *Birth Defects Res. Part A Clin. Mol. Teratol.* 82, 692–700 (2008).
53. FDA-DHHS. *Food Additives Permitted for Direct Addition to Food for Human Consumption; Follic Acid.* vol. 81 https://www.govinfo.gov/content/pkg/FR-2016-04-15/pdf/2016-08792.pdf (2016).
54. Aurova, G. A. et al. From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Carr. Protoc. Bioinform.* 43 (2013).
55. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum. Mutat.* 37, 233–245 (2016).
56. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443 (2020).
57. Thorvaldottir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* https://doi.org/10.1093/bib/bbs017 (2013).
58. Carbon, S. et al. The genome ontology resource: 20 years and still going strong. *Nucleic Acids Res.* 47, D330–D338 (2019).
59. Raudvere, U. et al. gProfiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198 (2019).
60. Firth, H. V. et al. DECIPHER: Database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* 84, 524–533 (2009).
61. Lohness, D. et al. Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants. *Development* 120, 2723–2748 (1994).
62. Willissey, H. R. et al. Katanin-like protein Katna2 is required for ciliogenesis and brain development in Xenopus embryos. *Dev. Biol.* 442, 276–287 (2018).
63. Alby, C. et al. Mutations in KIAA0586 cause lethal ciliopathies ranging from a hydrocephalus phenotype to short-rib polydactyly syndrome. *Am. J. Hum. Genet.* 97, 311–318 (2015).
Acknowledgements
We gratefully acknowledge the subjects of the study cohort for participation supported by program project 5P01HD035946-02, and the studies and participants who provide biological samples and data to gnomAD. Also, we acknowledge Sarah Riosa for technical support, and Tina Findley, Laura Farach and Parker Laville for their help to review and comment on the manuscript. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Institute of Child Health and Human Development; the National Institutes of Health; or the U.S. Department of Health and Human Services. Grant support: NIH NICHD R01HD073434.

Author contributions
K.S.A., H.N., and J.E.H. were responsible for study design, analyzing and reviewing data. K.S.A., H.N., A.C.M. and J.E.H. were responsible for monitoring experimental progress. A.A-K., K.S., and M.G. provided list of genes expressed in human neural tubes at CS12 and CS13. D-K.K. performed WES sequencing experiments and raw data QC. P.H., C.B., L.H., M.R.B. provided bioinformatics support on exome sequence data QC, filtering, and annotating variants. S.L. and J.G. were responsible for resequencing a subset of 88 subjects in study cohort and provided variant data for comparison. All authors were responsible for manuscript review.

Funding
NIH/NICHD R01HD073434.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-83058-7.

Correspondence and requests for materials should be addressed to K.S.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021