Human Genetics (2019) 138:703–713
https://doi.org/10.1007/s00439-019-02015-7

ORIGINAL INVESTIGATION

Exome sequencing of cases with neural tube defects identifies candidate genes involved in one-carbon/vitamin B12 metabolisms and Sonic Hedgehog pathway

Emeline Renard[1,2] · Céline Chéry[1,3,4] · Abderrahim Oussalah[1,3,4] · Thomas Josse[3,4] · Pascal Perrin[3,4] · Denise Tramoy[3,4] · Jimmy Voirin[1] · Olivier Klein[5] · Bruno Leheup[1,2] · François Feillet[1,2] · Rosa-Maria Guéant-Rodriguez[1,3,4] · Jean-Louis Guéant[1,3,4]

Received: 18 December 2018 / Accepted: 13 April 2019 / Published online: 28 May 2019

Abstract

Neural tube defects (NTD) result from complex mechanisms between genes, nutrition and environment. The identification of genetic predictors by genome exome sequencing and their influence on genome methylation need further consideration. Gene variants related to 1-CM metabolism (1-CM) could influence the methylation of genes involved in neural tube embryogenesis through impaired synthesis of *S*-adenosyl methionine. We performed exome sequencing of 6116 genes referenced in OMIM and NTD risk and genome-wide methylation in 23 NTD cases. We replicated the most significant associations in 81 other cases. The analysis of exome sequencing identified one gene of 1-CM, LRP2, and one gene of Sonic Hedgehog (SHH), *GLI3,* in the 23 NTD cases. The analysis restricted to genes of 1-CM and neural tube embryogenesis identified five gene predictors of 1-CM (*LRP2*, rs137983840; *MMAA*, rs148142853; *TCN2*, rs35838082; *FPGS*, rs41306702; *BHMT*, rs763726268) and two of SHH (*GLI3*, rs35364414; *MKS1*, rs151023718). We replicated the association of *TCN2*, *BHMT* and *GLI3* with NTD risk in the 81 cases. We found a significant hemimethylation of *CFAP46* that may influence SHH activation in one case, who carried risk alleles in *BHMT*, *LRP2*, *MMAA* and *GLI3*. In conclusion, we identified new candidate genes and rare variants that highlight an interacting influence of genes involved in SHH and 1-CM in the puzzle of genetic components of NTD risk.

Emeline Renard, Céline Chéry and Abderrahim Oussalah are equal contributors.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00439-019-02015-7) contains supplementary material, which is available to authorized users.

✉ Jean-Louis Guéant
jean-louis.gueant@univ-lorraine.fr

1 INSERM U1256, NGERE-Nutrition, Genetics, and Environmental Risk Exposure, Faculty of Medicine of Nancy, University of Lorraine, 54000 Nancy, Vandoeuvre-lès-Nancy, France

2 Department of Pediatrics, University Hospital of Nancy, 54000 Nancy, Vandoeuvre-lès-Nancy, France

3 Department of Molecular Medicine and Personalized Therapeutics, University Hospital of Nancy, 54000 Nancy, Vandoeuvre-lès-Nancy, France

4 Department of Biochemistry, Molecular Biology, Nutrition, and Metabolism, University Hospital of Nancy, 54000 Nancy, Vandoeuvre-lès-Nancy, France

5 Department of Pediatric Surgery, University Hospital of Nancy, 54000 Nancy, Vandoeuvre-lès-Nancy, France

Introduction

Neural tube defects are one of the most common congenital malformations, affecting approximately 0.5–1 child per 1000 births in Europe. These malformations are a consequence of the failure of fusion of the neural tube during the early stage of embryogenesis. They are classified in function of their localization along the central nervous system. The clinical consequence is variable, with lethal forms such as anencephaly or craniorachischisis, and forms with motor and sensitive deficits (as spina bifida aperta) and spina bifida occulta.

It is generally admitted that the pathomechanisms of neural tube defects (NTD) may result from complex interactions between genetic factors involved in neural tube development and metabolism, and environmental factors such as maternal nutritional status and intake of teratogenic drugs. A great variation of the prevalence of NTD is reported among populations worldwide. This may reflect differences in exposure to environmental risk factors or allele frequencies of genetic variants at specific loci.

The environmental and nutritional maternal risk factors of neural tube defect include age, folate status, caffeine consumption and anti-epileptic treatments by valproic acid and carbamazepine (Au et al. 2010). Since the 1980s, many studies have demonstrated the protective effect of perinatal folic acid fortification. This led to introducing the food fortification with folic acid in North America two decades ago and to recommend maternal acid folic supplementation 4 weeks before conception and during the first 12 weeks of pregnancy in Europe (1991). The fortification was associated with a large reduction of incidence of NTD worldwide (Crider et al. 2011). More recently, the contributing role of vitamin B12, another micronutrient involved in the one-carbon metabolism (1-CM), has also been pointed out (Guéant-Rodriguez et al. 2003; Senousy et al. 2018). Several mechanisms may underlie the association of 1-CM with NTD risk (Greene et al. 2009). One hypothesis is that the dysregulation of 1-CM decreases the intracellular production of *S*-adenosylmethionine (SAM), the universal methyl donor, with the consequences on DNA methylation and the subsequent epigenomic dysregulation of genes involved in neural tube embryogenesis.

The multifaceted involvement of genetic factors in NTD is mostly illustrated by the transgenic KO mice models of birth defects and the identification of several genetic polymorphisms in population studies. Many of these genes are involved in 1-CM and key pathways of the early stage of embryogenesis such as planar cell polarity (PCP) and WNT and Sonic Hedgehog (SHH) pathways (Greene et al. 2009). The genetic predictors for NTD risk in relation with 1-CM include the c.667C>T single nucleotide polymorphism (SNP) of *MTHFR*, the gene needed for the synthesis of methyltetrahydrofolate (MeTHF) (van der Put et al. 1998; Botto and Yang 2000), and other SNPs in other genes such as *MTR*, *MTRR*, *TCN2*, and *SCL19A1* (Guéant-Rodriguez et al. 2003; Candito et al. 2008; Afman et al. 2002; Guéant et al. 2003). Very few studies have used the next-generation sequencing (NGS) of the exome to identify less frequent and/or non-reported gene variants. In addition, the interplay of the potential genetic variants with DNA methylation has not received attention in the setting of NTD.

We made a systematic search of genetic variants referenced in OMIM and their gene ontology in the risk prediction of NTD. We investigated also that NTD risk could result in part from epigenome-wide alterations influenced by these genetic predictors. To address this issue, we performed a high-throughput exome sequencing using a large panel of 6116 genes and an epigenome-wide DNA methylome study in 23 children suffering from NTD. We replicated the most significant associations in 81 other NTD cases.

Methods

Study design

We performed a discovery study of mutations and rare genetic variants identified by NGS in 23 French NTD cases and a replication study in 81 other cases. We compared the alternative allele frequencies of cases with those from the ExAC database for the European non-Finnish population.

Study populations

We recruited 23 patients (10 males and 13 females) with myelomeningocele in the consultation of neurosurgery pediatric in Children Hospital of Nancy (France) between 2007 and 2008. Parental and patient consent was obtained after oral and written information. The study obtained the authorization of the ethical committee Est III (CPP) in 2007. The replication study was performed in a population previously described of 54 French and Italian NTD cases (Guéant-Rodriguez et al. 2003; Candito et al. 2008), including 48 cases with myelomeningocele and 6 cases with meningocele and 27 aborted French fetuses with myelomeningocele recruited in the Center of Foetopathology of the University Hospital Center of Nancy, France. We followed the principles outlined in the Helsinki Declaration, and we obtained informed and signed consents from the parents to perform the biochemical DNA analyses from the cord blood DNA, in the two replication populations.

NGS and Sanger sequencing

Genomic DNA from EDTA-treated peripheral blood samples was isolated using the *Nucleon*BACC3 Kit. (GE Healthcare, Aulnay-sous-Bois, France). The purified DNA samples were checked by 1% agarose electrophoresis and quantified by the PicoGreen dye binding assay. DNA sequencing was performed with a MiSeq® sequencer, from Illumina (San Diego, Ca, USA). We used the SureSelect Focused Exome Plus from Agilent® (Santa Clara, Ca, USA), which covers exonic regions of more than 6110 genes. We used the SureDesign tool to customize it with six genes described in the literature of NTD, which were not initially present in the panel (*DVL1*, *FZD5*, *DVL3*, *SNAI2*, *SHROOM2*, *TSPYL2*). Sequencing reads were aligned to GRCh37 hg19 with program Burrows-Wheeler Aligner (BWA-MEM) and sequence alignment/map (SAM) files were generated. Identification of single nucleotide variant (SNV) was performed using SNPPET SNP Caller. SureCall and Annovar were used for SNV

annotation. We confirmed the mutations identified in NGS by Sanger sequencing of PCR products, using the Big Dye Terminator® kit. v 1.1 (Applied Biosystems, Foster, Ca, USA) and capillary electrophoresis (ABI3100®).

The first phase of our bioinformatic analysis of NGS data is described in supplementary Figure S1. All the genetic variants retrieved from genome sequencing analysis of the genes used in the SureSelect Focused Exome Plus (Agilent®) design were imported for each of the 23 NTD cases. We checked systematically all the IGV image of the bam file for the variants, which were considered in our study. None of them was related to copy number loss. We used the following strategy for filtering and prioritizing the genetic variants. Step #1: import all retrieved genetic variants retrieved in the 23 NTD cases. Step #2: retain only low-frequency and rare variants by excluding variants with a reported alternative allele frequency (AAFc) > 5% in the following databases: the Exome Aggregation Consortium (ExAC), the Genome Aggregation Database (gnomAD), or the Single Nucleotide Polymorphism database (dbSNP 149). The aim of step #2 is to retain rare (AAFc < 1%) and low-frequency (AAFc < 5%) variants which exhibit relatively large effects on disease risk such as those involved in severe monogenic diseases. Step #3: exclude intronic or synonymous variants. Step #4: exclude variants with a low prediction probability for their association with neural tube defect risk based on a Phevor1 score < 0.02 (≥ 97th percentile). The Phevor score was calculated according to the RefSeq Genes 105v2, NCBI gene source, using the following ontologies: Human Phenotype Ontology, Gene Ontology, and OMIM Phenotype Ontology, and the following terms related to the NTD phenotype: "neural tube defect", "spina bifida", and "abnormal neural tube morphology". Step #5: exclude variants with a minor allele frequency among NTD cases (MAF_{NTD}) < 4% (at least 2 alleles present among the 23 NTD subjects). Step #6: exclude variants without genotype–phenotype relationship documented in the literature and with a "Benign" or "Likely Benign" prediction according to the ACMG Classification. Step #7: exclude variants with an enrichment ratio < 3 (enrichment ratio = MAF_{NTD}/AAFc).

The second phase of our analysis was the search for candidate genes that belong to the same ontology groups as those found in phase #1. Two groups were considered, genes that belong to embryogenesis in relation with neural tube closure and genes of 1-CM. For this second phase of bioinformatic analysis of NGS, we defined therefore two *in silico* gene panels, the first contained the selected genes for 1-CM (supplementary data Table S1) and the second the candidate genes implicated in embryogenesis or from animal models (http://ntdwiki.wikispaces.com, supplementary data Table S2). We selected the non-synonymous SNV of the genes included in the two panels, according to the deleterious or/and damaging prediction with Sift and Polyphen II, an AAFc < 5% and an enrichment ratio > 1.5. The allelic frequencies were compared with data from ExAC (http://exac.broadinstitute.org/) (European non-Finnish frequencies). Fischer's exact test was used to compare the allele frequency of NTD with that of the reference population.

In the third phase of the study, we evaluated the degree of enrichment of the variants considered as potentially relevant in phases #1 and 2 of the study using two distinct populations of NTD patients originating from France and Italy ($n = 81$). In phase #3, Fischer's exact test was used to compare the MAF_{NTD} to the AAFc. Using the ExAC population ($n = 60{,}706$) as a reference population for retrieving AAFc, we performed a study power to estimate the minimal number of NTD subjects required for replicating the observed associations in phases #1 and #2 of the study. Using a hypothetical AAFc of 0.025, a MAF_{NTD} of 0.075 (enrichment ratio 3), a type I error probability associated with the Fisher's exact test for the null hypothesis of 0.05, and a study power of 0.80, the minimal number of experimental subjects necessary to achieve a statistical significance was 66 (132 alleles).

DNA methylome analysis

600 ng of blood DNA was bisulfite modified using EZ DNA Methylation kit (Zymo Research, Proteigene, Saint-Marcel, France). The genome-wide profiling of methylome was determined using the Infinium Human Methylation 450 BeadChip array (Illumina, Paris, France), according to the manufacturer's instructions. The arrays were scanned on an Illumina iScan® system, and raw methylation data were extracted using Illumina's Genome Studio methylation module. Methylation was described as a β value, ranging between 0 (fully unmethylated CpG) and 1 (fully methylated CpG). Background correction and normalization were implemented using the SWAN method (R Package Minfi) (Wu et al. 2014).

We performed a comparison of methylation levels using two sets of comparison: Set #1: patient(s) with several replicated rare SNV of 1-CM and Set #2: patients with replicated rare SNV of B12 metabolism. Set #1: a patient with *BHMT-LRP2-MMAA* and *GLI3* variants was identified; Set #2: two patients with the *TCN2* variant were identified (rs35838082). DNA methylome data from identified sets were compared with those from pseudo-controls, which consisted of NTD cases from the discovery cohort with no above-mentioned variants. Due to the low sample size, and considering the exploratory approach of our analysis, we used the smoothed *P* value transformation by converting nominal *P* values obtained from the *t* test to smoothed *P* values using a window radius of 3, as previously reported (Guéant et al. 2018). To assess the methylome architecture, we performed linkage disequilibrium (LD)-pairwise analysis on all adjacent CpG pairs in a chromosome. D' values were used in the epi-LD

plots. All epi-LD analyses were performed after transforming the CpG β values to categorical variables according to the ENCODE project, as previously reported (Guéant et al. 2018; ENCODE Project Consortium 2012). We estimated the haplotype frequencies using the expectation/maximization (EM) algorithm (Gabriel et al. 2002; Remington et al. 2001). We used the following parameters for the haplotype block definition: confidence bound: 0.7–0.98; reject criteria threshold (*D'*): 0.9; minor epi-allele frequency threshold: 0.05; Max # Markers in a Block: 30; max length of a block: 160; display threshold: 0.01. All statistical analyses were performed using the SNP and Variation Suite (v8.8.1; Golden Helix, Inc., Bozeman, MT, USA).

Results

Characteristics of the study populations

The 23 cases recruited in the discovery study and the postnatal cases of the replication study were aged between 14 and 35 years (mean 25 years) and 5 and 21 years (mean 8.5 years), respectively. The clinical characteristics of the discovery cohort are summarized in supplementary Table S3. All cases had myelomeningocele.

In the replication study, among the 54 NTD postnatal cases, 41 had myelomeningocele, and 13 cases had meningocele. All the 27 fetuses of the replication group had myelomeningocele.

Identification of NTD risk candidate genes in NGS analysis

The study design is summarized in Fig. 1. The genome exome analysis of the 23 NTD cases retrieved 55,798 genetic variants. Among them, 35,233 represented SNP with an allele frequency > 5% and were excluded (Supplementary data Figure S1). Among the low-frequency and rare variants retained, 15,015 were annotated as intronic or synonymous variants and were thus excluded. The Phevor score was calculated for the remaining 5541 variants and allowed to exclude variants exhibiting low proximity with the NTD phenotype (Phevor score < 0.02; < 97th percentile). Among the 178 remaining variants, 42 were retained given their $MAF_{NTD} > 4\%$. After excluding the variants with a "Benign" or "Likely Benign" prediction according to the ACMG Classification and those exhibiting an enrichment < 3, three variants from two genes were finally retained as potential candidates for the replication study (Table 1). These genes are involved in the neurodevelopment and ciliopathies in relation with NTD (*GLI3*) and 1-CM (*LRP2*).

According to these results, we restricted subsequently our search of gene predictors to the genes involved in the 1-CM and neurodevelopment/ciliopathies. We identified alleles of

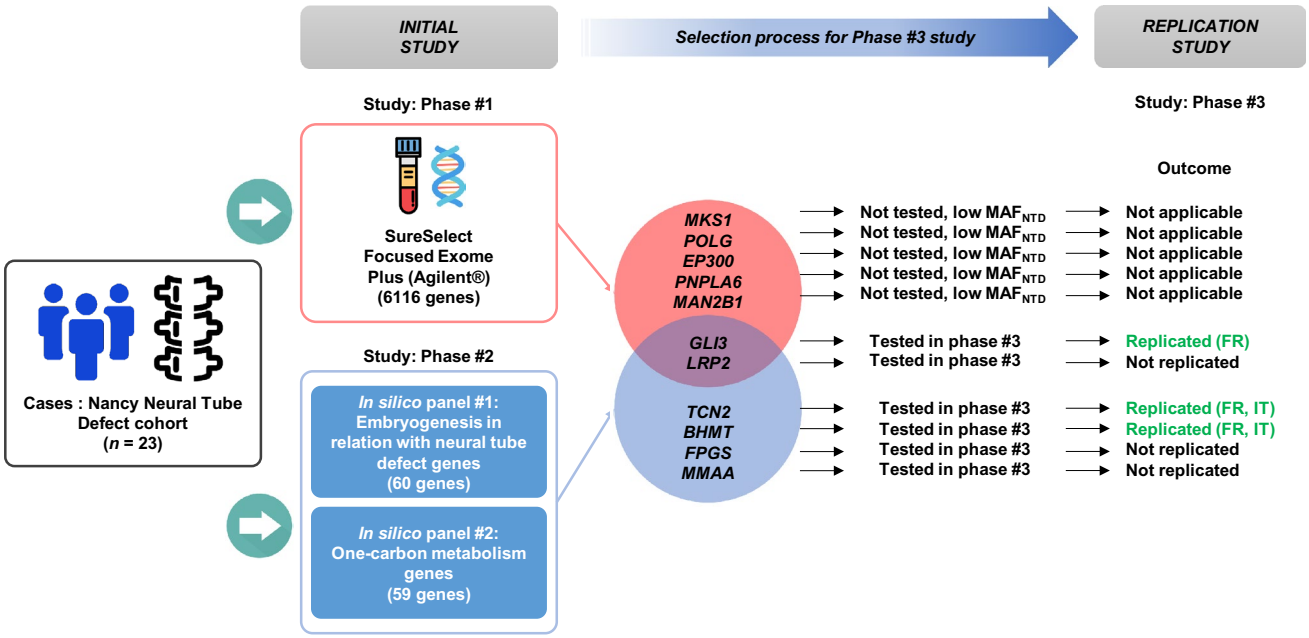


Fig. 1 Study design. The study was designed in three phases. The first phase analyzed the genetic variants retrieved from genome sequencing analysis in 23 cases with neural tube defects (NTD). In the second phase, we analyzed candidate genes that belonged to the same ontology groups as the genes identified in phase 1. In the third phase, we replicated the significant associations of candidate genes with NTD risk found in phase #1 and phase #2 for gene variants with an MAF higher than 0.01, in 81 other NTD cases including 41 French (FR) and 40 Italian (IT) cases. Abbreviation: Minor allele frequency in NTD (MAF_{NTD})

Table 1 Variants retrieved from the whole NGS analysis of the 23 NTD cases (phase #1)

Genetic variant	[Ref/Alt]	Gene names	HGVS nomenclature	Sequence ontology	MAF_{NTD}	AAFc	Enrichment (MAF_{NTD}/AAFc)	Considered for phase #3
7:42004062-SNV	G/A	*GLI3*	NM_000168.5:c.4609C>T NP_000159.3:p.Arg1537Cys	Missense	0.152	0.05	3.0	Yes
7:42003898-SNV	C/A	*GLI3*	NM_000168.5:c.*30G>T –	3_prime_UTR	0.152	0.05	3.0	Yes
2:170097707-SNV	T/G	*LRP2*	NM_004525.2:c.3836A>C NP_004516.2:p.Asp1279Ala	Missense	0.065	0.022	3.0	Yes
17:56290344-SNV	T/C	*MKS1*	NM_017777.3:c.857A>G NP_060247.2:p.Asp286Gly	Missense	0.043	0.0009	48.3	No
15:89873364-SNV	C/G	*POLG*	NM_002693.2:c.803G>C NP_002684.1:p.Gly268Ala	Missense	0.043	0.004	10.9	No
22:41546158-SNV	C/A	*EP300*	NM_001429.3:c.2773C>A NP_001420.2:p.Pro925Thr	Missense	0.043	0.006	7.2	No
19:7626428-SNV	G/A	*PNPLA6*	NM_001166111.1:c.4108G>A NP_001159583.1:p.Gly1370Ser	Missense	0.043	0.008	5.4	No
19:12757392-SNV	C/T	*MAN2B1*	NM_000528.3:c.*42G>A –	3_prime_UTR	0.043	0.009	4.8	No

HGVS human genome variation society, MAF_{NTD} minor allele frequency, *AAFc* alternative allele frequency in the following databases in reference populations (ExAC, gnomAD, or dbSNP 149)

16 genes from the panel of genes 1-CM and 5 genes from the panel of genes implicated during embryogenesis, with an AAFc < 5% (ExAC) and an enrichment ratio > 1.5 (Table 2). Among the genes of 1-CM, five were significantly associated with an increased risk of NTD. Three of these genes were involved in vitamin B12 metabolism, *LRP2* (rs137983840, $P = 0.005$), *MMAA* (rs148142853, $P = 0.005$) and *TCN2* (rs35838082, $P = 0.044$), one in folate cellular metabolism, *FPGS* (rs41306702, $P = 0.0012$) and one in choline metabolism, *BHMT* (rs763726268, $P = 0.011$) (Table 2). The *TCN2* rs35838082 was annotated as benign based on the ACMG classification criteria/ClinVar database. However, a recent article showed its association with a decreased plasma concentration of holotranscobalamin (Sobczyńska-Malefora et al. 2016). We also detected an alternative allele for the rs117687681 of *CBS* associated with NTD risk with borderline significance ($P = 0.06$). These genetic variants were predicted to be 'probably damaging' or 'damaging'. The alternative allele frequencies for the rs763726268, rs41306702, rs137983840, rs148142853, rs35838082 and rs117687681 were, respectively, enriched by 180.7-, 11.9-, 209.7-, 209.7-, 5.9- and 17.6-fold in the NTD population.

We replicated the association of *BHMT* rs763726268 and *TCN2* rs35838082 with NTD risk in the 81 other cases compared with the ExAC database (Table 3). One NTD case was homozygous for the *BHMT* rs763726268.

In the neural tube embryogenesis panel, we found variants from five genes with an AAFc < 5% (ExAC) and an enrichment ratio > 1.5. The SNV of these five genes are detailed in Table 2. Among them, only two variants were significantly associated with NTD risk, *GLI3* (c.4609C>T, rs35364414) and *MSK1* (c.857A>G, rs151023718). We reported the *MKS1* variant in two heterozygous cases and the *GLI3* variant in six cases, including one homozygous and five heterozygous. The *MKS1* allele frequency was at 4.3% in the NTD case. This variant is extremely rare in the European sample population from ExaC, with an AAF of 0.09% and could not be retained for the replication study. The *GLI3* c.4609T allele frequency of 15.2% was significantly higher than that reported in the European non-Finish population (4.7% in ExAC data, $P = 0.012$). This variant is predicted as deleterious and probably damaging by SIFT and Polyphen II, respectively. It is in LD with another variant c.4007G>A, (rs35280470) of *GLI3*. We did not replicate the association of *GLI3* rs35364414 in the 81 NTD cases (Table 3). However, the allelic frequency of rs35364414 was 12.2% in the 41 French NTD vs. 4.7% in ExAC ($P = 0.005$).

One patient presented four of the rare variants identified in the discovery population at a heterozygous status: *BHMT* rs763726268, *LRP2* rs137983840, *MMAA* rs148142853 (from the 1-CM panel), and *GLI3* rs35364414, from the neural tube embryogenesis panel (case NTD4, Sanger sequencing presented in supplementary Figure S1).

Methylation analysis

We hypothesized that gene variants related to 1-CM metabolism could influence the methylome through impaired synthesis of SAM. We performed the genome profiling of methylation in 22 of the 23 NTD cases (12 females and 10 males) and focused our interest on those who had the presence of one or several replicated variants. We found a top significant signal in the vicinity of the *CFAP46* gene (alias, *TTC40*) in patient S27 who carried risk alleles in *BHMT*, *LRP2*, *MMAA,* and *GLI3*, as shown in Fig. 2. This top significant locus encompassed one gene (*CFAP46*)

Table 2 Genetic variants reporting low-frequency variants (<5%) in the NTD cases with an enrichment of at least 1.5-fold in comparison with the ExAC database (European non-Finnish) retrieved from the 1-CM panel (OCM) and neural tube embryogenesis panel (NTE) (#phase 2)

Gene	Panel	Variant	HGVS, cDNA	HGVS, Prot	AAC_{NTD}	MAF_{NTD}	AAFc	Enrichment (NTD vs. ExAC)	SIFT	Polyphen2	*P* value
AHCY	OCM	rs41301825	c.367G>A	p.Gly123Arg	1	0.022	6.24×10^{-3}	3.5	D	B	0.25
BHMT	OCM	rs763726268	c.1196delA	p.Lys400AsnfsTer15	1	0.043	2.38×10^{-4}[a]	180.7	–	–	0.011
CBS	OCM	rs117687681	c.1105C>T	p.Arg369Cys	2	0.043	2.44×10^{-3}	17.6	D	P	0.06
DNMT1	OCM	Not reported	c.3830G>A	p.Arg1277Gln	1	0.022	Not reported	–	D	D	–
FPGS	OCM	rs41306702	c.253C>T	p.Arg85Trp	2	0.043	3.61×10^{-3}	11.9	D	P	0.0012
GIF	OCM	rs150884181	c.290T>C	p.Met97Thr	1	0.022	1.12×10^{-2}	2.0	D	P	0.4
LRP2	OCM	rs149469954	c.5209C>T	p.Leu1737Phe	1	0.022	1.70×10^{-3}	13.0	D	B	0.07
LRP2	OCM	rs137983840	c.10030G>A	p.Ala3344Thr	1	0.022	1.05×10^{-4}	209.7	D	P	0.005
MMAA	OCM	rs148142853	c.941G>A	p.Arg314His	1	0.022	1.05×10^{-4}	209.7	D	B	0.005
MTFMT	OCM	rs199599204	c.16C>T	p.Arg6Trp	1	0.022	1.32×10^{-2}	1.7	D	B	0.46
MTHFD1	OCM	Not reported	c.1745T>G	p.Met582Arg	2	0.043	Not reported	–	D	P	–
MTR	OCM	rs116836001	c.3079C>T	p.Arg1027Trp	1	0.022	3.60×10^{-3}	6.1	D	B	0.15
MTR	OCM	rs12749581	c.155G>A	p.Arg52Gln	1	0.022	5.90×10^{-3}	3.7	D	B	0.23
MTRR	OCM	rs41283145	c.1549A>G	p.Thr517Ala	1	0.022	6.46×10^{-3}	3.4	D	P	0.26
PEMT	OCM	Not reported	c.311G>A	p.Arg104His	1	0.022	Not reported	–	D	P	–
TCN2	OCM	rs35838082	c.643C>T	p.Arg215Trp	2	0.043	7.24×10^{-3}	5.9	D	D	0.044
BRCA1	NTE	rs1800744	c.4598G>T	p.Ser1533Ile	1	0.022	3.18×10^{-3}	6.9	D	B	0.14
GLI2	NTE	rs114814747	c.4558G>A	p.Asp1520Asn	1	0.022	1.41×10^{-2}	1.6	D	D	1
GLI3	NTE	rs35364414	c.4609C>T	p.Arg1537Cys	7	0.152	4.73×10^{-2}	3.2	D	D	0.012
HK2	NTE	rs146476722	c.2207C>T	p.Pro736Leu	1	0.022	1.96×10^{-3}	11.2	D	D	0.09
MKS1	NTE	rs151023718	c.857A>G	p.Asp286Gly	2	0.043	9×10^{-4}	48.3	D	P	0.0008

AAC alternative allelic count, MAF_{NTD} minor allele frequency in NTD, *AAFc* alternative allele frequencies from the ExAC Database (European non-Finnish), *D* deleterious, *T* tolerated, *B* benign, *P* probably damaging, *D* damaging

[a]Data were retrieved from the gnomAD database

Table 3 Results of replication study for the seven variants identified in the study in the 81 cases of NTD and comparison with alternative allelic frequency reported in the ExAC database

	HGVS, cDNA	HGVS, Prot	MAF_{NTD}	AAFc	*P* value[a]
BHMT rs763726268	c.1196delA	p.Lys400AsnfsTer15	0.012	2.706×10^{-4}	0.001
TCN2 rs35838082	c.643C>T	p.Arg215Trp	0.037	7.239×10^{-3}	0.001
MMAA rs148142853	c.941G>A	p.Arg314His	0.000	0.000	1
LRP2 rs137983840	c.10030G>A	p.Ala3344Thr	0.000	0.000	1
FPGS rs41306702	c.253C>T	p.Arg85Trp	0.000	3.605×10^{-3}	1
GLI3 rs35364414	c.4609C>T	p.Arg1537Cys	0.068	0.047	0.195

MAF_{NTD} minor allele frequency, *AAFc* alternative allele frequency from ExAC database (European non-Finnish)

[a]Fisher's exact test

and two long intergenic non-protein coding RNA, namely: *LINC01166* (alias, *LOC101927590*), and *LINC01168* (alias, *LOC399829*) (Fig. 2a, b). The CpG probes associated with the top *P* values were: cg01263624, cg00912926, cg23987897, and cg17693826. They were hemimethylated in the S27 case, while they were fully unmethylated in the NTD pseudo-controls (Fig. 2c and Supplementary Table S4). We performed an epi-haplotype approach for estimating the methylome architecture of this locus and found a relatively structured co-methylation pattern encompassing *CFAP46*, *LINC01166*, and *LINC01168* (Fig. 2b).

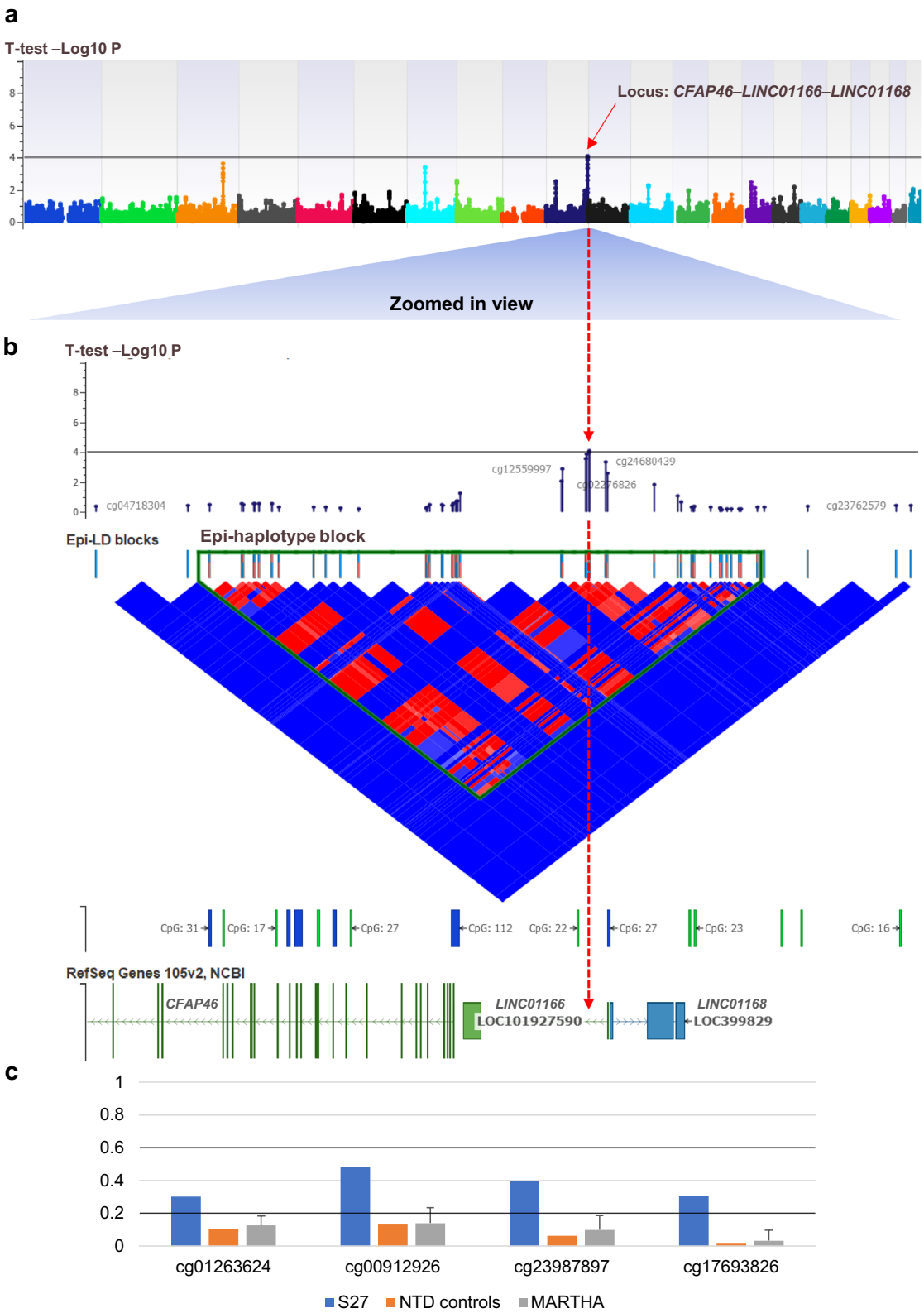
Discussion

Our study of the genome sequencing of a large panel of genes has permitted to identify new variants associated with NTD risk in genes involved in vitamin B12 metabolism (*LRP2*, *MMAA*, and *TCN2*), folate cellular metabolism (*FPGS*), choline metabolism (*BHMT*), and one gene involved in the Sonic Hedgehog pathway (*GLI3*).

We found three variants, which predicted NTD risk in genes of B12 metabolism, *LRP2*, *TCN2* and *MMAA,* in the discovery cohort. Among them, the *TCN2* variant was the single to be significantly associated with NTD risk in the replication study. We found a risk prediction of the rs35838082 missense variant c.643C>T (p.Arg215Trp) of *TCN2*, the gene that encodes transcobalamin. This protein plays a key role in B12 transport and cell delivery (Green et al. 2017). It is internalized by receptor-mediated endocytosis through its binding to CD320 in the brain and low-density lipoprotein (LDL) receptor-related protein 2 (LRP2) in the kidney tubule and other tissues. Whether this polymorphism influences the binding of transcobalamin to either CD320 or LRP2 should deserve further attention. Other variants of *TCN2* have been previously associated with NTD, including *TCN2* c.776C>G in the Indian population with a high prevalence of B12 deficiency (Godbole et al. 2011). We found a significant association of *LRP2* rs137983840 *c.10030G>A* with NTD risk. LRP2, also named megalin, is a multifunctional endocytic receptor expressed on the apical surface of the neuroepithelium in the early steps of embryonic development (Willnow et al. 2007). Previous studies have reported other variants of *LRP2* gene in association with NTD risk (Prasoona et al. 2018). The loss of *LRP2* expression in the central nervous system leads to impaired closure of the rostral neural tube in mice (Kur et al. 2014). The association of *LRP2* with NTD risk could be related to its role in folate and vitamin B12 uptake through its binding to *FOLR1* and cubilin, respectively (Kur et al. 2014). The involvement of the *MMAA* gene in the pathomechanisms of NTD is less clear and not documented in the literature. *MMAA* plays a key role in the transport of vitamin B12 in the mitochondrion and mutations in this gene increase methylmalonic acid (Green et al. 2017). High levels of MMA are associated with NTD (Senousy et al. 2018).

The variant identified in *BHMT* is rare and has not been reported previously. It is now referenced as rs763726268 and has a very low allele frequency in the general population (Table 2). The deletion leads to loss of a stop codon, responsible for an elongation of the protein. *BHMT* encodes betaine-homocysteine *S*-methyltransferase, a protein, which catalyzes homocysteine and betaine transformation in methionine. It is mainly expressed in liver and kidney in humans (Lee et al. 2012). The association of other variants of *BHMT* gene with NTD risk has been previously described in the literature (Cao et al. 2018; Morin et al. 2003). Moreover, low maternal choline status produces a potential risk for NTD occurrence in some, but not all published studies (Shaw et al. 2009; Mills et al. 2014).

Our study is the first report of NTD association with *GLI3*. The rs35364414 was a risk predictor in the discovery study and in the French cases of the replication study. GLI3 is a transcriptional factor involved in the Sonic Hedgehog pathway (SHH). The SHH pathway is essential during development, particularly for the formation and patterning of the neural tube. The loss of SHH signaling is associated with a cranial defect such as holoprosencephaly and cyclopia. Conversely, hyperactivation of the SHH pathway seems to be associated with neural tube defect, as observed



notably with mutations of *PATCH1* (Murdoch and Copp 2010). *GLI3* rs35364414 is a missense *c.4609C>T* variant (p.Arg1537Cys) located in an exon needed for the translation of the C-terminal domain. This domain is phosphorylated by the protein kinase A to obtain the truncated and repressive form named GLI3R. It has been hypothesized that the

◂**Fig. 2** DNA methylation. The DNA methylation profiling of patient S27 exhibiting genetic variants in *BHMT*, *LRP2*, and *MMAA* in comparison with pseudo-controls (seven NTD cases with that did not exhibit the 1-CM variants). **a** Top significant signal in the vicinity of the *CFAP46* gene *(alias TTC40)*; **b** epi-haplotype approach for estimating the methylome architecture of this locus showing a relatively structured co-methylation pattern encompassing *CFAP46*, *LINC01166*, and *LINC01168*; **c** beta values of the four CpG probes associated with a top statistical significance in the DNA methylome analysis between the case S27 and pseudo-controls (MARTHA cohort), as described previously

reduced GLI3R is the fundamental cause of the failure of neural tube closure, but this hypothesis was challenged by the absence of risk association with NTD (Murdoch and Copp 2010). GLI3 shares similarities with GLI2, another transcriptional factor involved in NTD pathomechanisms. Pan et al. have previously shown that a *Gli2* mutation in the site recognized by PKA (GLI P1-4) produces NTD in mice (Pan et al. 2009). We found also two heterozygous cases for rs151023718 of *MSK1*, a gene was involved in SHH and ciliary function. It belongs to the syndromic encephalocele genes associated with Bardet–Biedl syndrome (Leitch et al. 2008). We could not replicate the association of rs151023718 with NTD as the MAF is extremely rare in the European population.

Our and previous data suggest interactions between 1-CM and SHH. A link between 1-CM and *GLI3* was evidenced by Ernest et al. (Ernest et al. 2002), who showed that mice mutant *Gli3xtj* (heterozygote) has increased level of homocysteine compared with wild-type mice. Homozygous *Gli3xtj* mice were not viable and presented anencephaly. In our cohort, the homocysteine level in patients with *GLI3 c.4609C>T* variant was 13.5 µmol/L vs. 10.5 µmol/L in cases with no *GLI3* mutation ($P=0.14$). Recently, Toriyama et al. showed that the folate-dependent methylation of septins governs ciliogenesis during neural tube closure (Toriyama et al. 2017). Septin-2 is methylated in two residues and this methylation regulates septin-2 functions in animal models. Moreover, the hypomethylation produced by folate depletion and Adox treatment influence SHH signaling by decreasing the truncated Gli3R in Scl19a1$^{-/-}$ cells (Ernest et al. 2002). The impaired Gli3 processing is expected to activate the SHH pathway. One patient carried four variants encompassing *LRP2*, *BHMT*, *MMAA*, and *GLI3*, and which were individually identified as NTD predictors. The interacting and complementary role of these gene variants may reflect epistasis (Mackay and Moore 2014) and are consistent with the "omnigenic" hypothesis of the genetic predictors of NTD (Boyle et al. 2017; Chen et al. 2018). LRP2 is described as a component of SHH in neural tube development by influencing the internalization and cellular trafficking of SHH/patched 1 complex (Christ et al. 2012; McCarthy et al. 2002). The Gli3 processing could be amplified by a decreased methylation capacity produced by the *BHMT* variant (Murdoch and Copp 2010). The epigenome-wide analysis identified a hemimethylated locus encompassing the *CFAP46* gene and two long intergenic non-protein coding RNA, *LINC01166* (alias *LOC101927590*), and *LINC01168* (alias *LOC399829*). *LINC01166* encodes a long non-coding RNA, which is almost exclusively expressed in the brain and testis (3.9 TPM) according to data extracted from the Genotype-Tissue Expression Project (GTEx) (gtexportal.org) (Supplementary data Figure S1). The *CFAP46* gene is expected to encode a "protein cilia and flagella-associated protein 46" involved in cilium axoneme and cilium movement. The primary cilium is abundant in the neural tube. Most of the components of the SHH pathway, such as PTCH1 (Goodrich et al. 1997), Gli2 and Gli3 proteins (Haycraft et al. 2005), are localized in the primary cilium (Pal and Mukhopadhyay 2015). Mutations of proteins required for proper functioning of cilia and SHH signaling can disrupt neural tube closure (Murdoch and Copp 2010). Moreover, *CFAP46* is involved in pathomechanisms of heterotaxy syndrome (Liu et al. 2018), a syndrome with an increased risk of NTD (Loomba et al. 2015).

We hypothesized that part of NTD risk could result from methylome changes of genes involved in NTD pathogenesis. These methylation alterations could be a consequence of the disruption of 1-CM homeostasis and production of SAM by nutritional, metabolic, and gene determinants that potentially influence DNA methylation in tissues, as previously reported (Guéant et al. 2013; Ingrosso et al. 2003). The relation between 1-CM and DNA methylation could explain in part the protective effect of folate on NTD. Similarly, a homocysteine-dependent shift from monoallelic to biallelic expression of imprinting genes is reverted by folate treatment, in PBMC from patients with uremia (Ingrosso et al. 2003). Presently we found only one methylome change in a locus encompassing *CFAP46*, *LINC01166*, and *LINC01168*, among genes involved in neural tube embryogenesis, in blood DNA from cases carrying NTD gene predictors related to 1-CM. We assume that more significant changes would be expected in ectoderm tissue studied at the time window of neural tube closure.

Our study has several limitations. The population of the discovery study had a very limited size. To overcome this difficulty, we carried out a replication study that confirmed the NTD risk associations for three of the variants identified as predictors in the discovery study. The second limitation is the fact that we had no parental genetic data available to study the transmission of risk alleles. The third limitation is that we performed a replication study in a population combining NTD cases with DNA collected in either fetopathology samples or after birth.

In conclusion, we performed an NGS study that identified new candidate genes and new rare variants with risk

prediction of NTD. These variants were located in interacting genes that may reflect the influence of deregulated B12 metabolism and SHH pathway in the pathomechanisms of NTD.

Author contributors ER: conducted research; analyzed data; performed genotyping analyses; performed statistical analysis; wrote paper. CC: conducted research; analyzed data; performed genotyping analyses. AO: conducted research; analyzed data; performed statistical analysis; wrote paper. TJ: conducted research; performed genotyping analyses. JV and OK: recruited subjects. FF and BL: conducted research; analyzed data; recruited subjects. R-MG-R: conducted research; provided essential materials; participated in genotyping analyses; performed statistical analysis; recruited subjects, analyzed data. J-LG: designed and coordinated research; analyzed data; wrote paper; had primary responsibility for final content.

Funding This work was supported partly by the French PIA project "Lorraine Université d'Excellence", reference ANR-15-IDEX-04-LUE

Compliance with ethical standards

Conflict of interest None declared.

Ethics approval The studies described here were approved by the Nancy University Hospital ethics committee.

References

Afman LA, Lievers KJA, van der Put NMJ et al (2002) Single nucleotide polymorphisms in the transcobalamin gene: relationship with transcobalamin concentrations and risk for neural tube defects. Eur J Hum Genet EJHG 10:433–438. https://doi.org/10.1038/sj.ejhg.5200830

Au KS, Ashley-Koch A, Northrup H (2010) Epidemiologic and genetic aspects of spina bifida and other neural tube defects. Dev Disabil Res Rev 16:6–15. https://doi.org/10.1002/ddrr.93

Botto LD, Yang Q (2000) 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. Am J Epidemiol 151:862–877

Boyle EA, Li Y, Pritchard JK (2017) An expanded view of complex traits: from polygenic to omnigenic. Cell 169:1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Candito M, Rivet R, Herbeth B et al (2008) Nutritional and genetic determinants of vitamin B and homocysteine metabolisms in neural tube defects: a multicenter case-control study. Am J Med Genet A 146A:1128–1133. https://doi.org/10.1002/ajmg.a.32199

Cao L, Wang Y, Zhang R et al (2018) Association of neural tube defects with gene polymorphisms in one-carbon metabolic pathway. Childs Nerv Syst ChNS Off J Int Soc Pediatr Neurosurg 34:277–284. https://doi.org/10.1007/s00381-017-3558-z

Chen Z, Lei Y, Zheng Y, Aguiar-Pulido V et al (2018) Threshold for neural tube defect risk by accumulated singleton loss-of-function variants. Cell Res 28:1039–1041. https://doi.org/10.1038/s41422-018-0061-3

Christ A, Christa A, Kur E et al (2012) LRP2 is an auxiliary SHH receptor required to condition the forebrain ventral midline for inductive signals. Dev Cell 22:268–278. https://doi.org/10.1016/j.devcel.2011.11.023

Crider KS, Bailey LB, Berry RJ (2011) Folic acid food fortification-its history, effect, concerns, and future directions. Nutrients 3:370–384. https://doi.org/10.3390/nu3030370

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. https://doi.org/10.1038/nature11247

Ernest S, Christensen B, Gilfix BM et al (2002) Genetic and molecular control of folate-homocysteine metabolism in mutant mice. Mamm Genome 13:259–267. https://doi.org/10.1007/s00335-001-3054-2

Gabriel SB, Schaffner SF, Nguyen H (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229

Godbole K, Gayathri P, Ghule S et al (2011) Maternal one-carbon metabolism, MTHFR and TCN2 genotypes and neural tube defects in India. Birt Defects Res A Clin Mol Teratol 91:848–856. https://doi.org/10.1002/bdra.20841

Goodrich LV, Milenković L, Higgins KM, Scott MP (1997) Altered neural cell fates and medulloblastoma in mouse patched mutants. Science 277:1109–1113

Green R, Allen LH, Bjørke-Monsen A-L et al (2017) Vitamin B12 deficiency. Nat Rev Dis Primer 3:17040. https://doi.org/10.1038/nrdp.2017.40

Greene NDE, Stanier P, Copp AJ (2009) Genetics of human neural tube defects. Hum Mol Genet 18:R113–R129. https://doi.org/10.1093/hmg/ddp347

Guéant J-L, Guéant-Rodriguez R-M, Anello G et al (2003) Genetic determinants of folate and vitamin B12 metabolism: a common pathway in neural tube defect and Down syndrome? Clin Chem Lab Med 41:1473–1477. https://doi.org/10.1515/CCLM.2003.226

Guéant JL, Namour F, Guéant-Rodriguez RM, Daval JL (2013) Folate and fetal programming: a play in epigenomics? Trends Endocrinol Metab 24(6):279–289. https://doi.org/10.1016/j.tem.2013.01.010

Guéant J-L, Chéry C, Oussalah A et al (2018) APRDX1 mutant allele causes a MMACHC secondary epimutation in cblC patients. Nat Commun 9:67. https://doi.org/10.1038/s41467-017-02306-5

Guéant-Rodriguez RM, Rendeli C, Namour B et al (2003) Transcobalamin and methionine synthase reductase mutated polymorphisms aggravate the risk of neural tube defects in humans. Neurosci Lett 344:189–192

Haycraft CJ, Banizs B, Aydin-Son Y et al (2005) Gli2 and Gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function. PLoS Genet 1:e53. https://doi.org/10.1371/journal.pgen.0010053

Ingrosso D, Cimmino A, Perna AF et al (2003) Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia. Lancet 361(9370):1693–1699

Kur E, Mecklenburg N, Cabrera RM et al (2014) LRP2 mediates folate uptake in the developing neural tube. J Cell Sci 127:2261–2268. https://doi.org/10.1242/jcs.140145

Lee MB, Kooistra M, Zhang B et al (2012) Betaine homocysteine methyltransferase is active in the mouse blastocyst and promotes inner cell mass development. J Biol Chem 287:33094–33103. https://doi.org/10.1074/jbc.M112.365478

Leitch CC, Zaghloul NA, Davis EE et al (2008) Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet–Biedl syndrome. Nat Genet 40:443–448. https://doi.org/10.1038/ng.97

Liu C, Cao R, Xu Y et al (2018) Rare copy number variants analysis identifies novel candidate genes in heterotaxy syndrome patients with congenital heart defects. Genome Med 10:40. https://doi.org/10.1186/s13073-018-0549-y

Loomba R, Shah PH, Anderson RH (2015) Fetal magnetic resonance imaging of malformations associated with heterotaxy. Cureus 7:e269. https://doi.org/10.7759/cureus.269

Mackay TF, Moore JH (2014) Why epistasis is important for tackling complex human disease genetics. Genome Med 6:124. https://doi.org/10.1186/gm561

McCarthy RA, Barth JL, Chintalapudi MR et al (2002) Megalin functions as an endocytic sonic hedgehog receptor. J Biol Chem 277:25660–25667. https://doi.org/10.1074/jbc.M201933200

Mills JL, Fan R, Brody LC et al (2014) Maternal choline concentrations during pregnancy and choline-related genetic variants as risk factors for neural tube defects. Am J Clin Nutr 100:1069–1074. https://doi.org/10.3945/ajcn.113.079319

Morin I, Platt R, Weisberg I et al (2003) Common variant in betaine-homocysteine methyltransferase (BHMT) and risk for spina bifida. Am J Med Genet A 119A:172–176. https://doi.org/10.1002/ajmg.a.20115

MRC Vitamin Study Research Group (1991) Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. Lancet Lond Engl 338:131–137

Murdoch JN, Copp AJ (2010) The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. Birt Defects Res A Clin Mol Teratol 88:633–652. https://doi.org/10.1002/bdra.20686

Pal K, Mukhopadhyay S (2015) Primary cilium and sonic hedgehog signaling during neural tube patterning: role of GPCRs and second messengers. Dev Neurobiol 75:337–348. https://doi.org/10.1002/dneu.22193

Pan Y et al (2009) Phosphorylation of Gli2 by protein kinase A is required for Gli2 processing and degradation and the Sonic Hedgehog-regulated mouse development. Dev Biol 1:177–189

Prasoona R, Sunitha T, Srinadh B et al (2018) LRP2 gene variants and their haplotypes strongly influence the risk of developing neural tube defects in the fetus: a family-triad study from South India. Metab Brain Dis 33:1343–1352. https://doi.org/10.1007/s11011-018-0242-2

Remington DL, Thornsberry JM, Matsuoka Y et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484. https://doi.org/10.1073/pnas.201394398

Senousy SM, Farag MK, Gouda AS et al (2018) Association between biomarkers of vitamin B12 status and the risk of neural tube defects. J Obstet Gynaecol Res 44:1902–1908. https://doi.org/10.1111/jog.13751

Shaw GM, Finnell RH, Blom HJ et al (2009) Choline and risk of neural tube defects in a folate-fortified population. Epidemiol Camb Mass 20:714–719. https://doi.org/10.1097/EDE.0b013e3181ac9fe7

Sobczyńska-Malefora A, Pangilinan F, Plant GT et al (2016) Association of a transcobalamin II genetic variant with falsely low results for the holotranscobalamin immunoassay. Eur J Clin Invest 46:434–439. https://doi.org/10.1111/eci.12617

Toriyama M, Toriyama M, Wallingford JB, Finnell RH (2017) Folate-dependent methylation of septins governs ciliogenesis during neural tube closure. FASEB J 31:3622–3635. https://doi.org/10.1096/fj.201700092R

van der Put NM, Gabreëls F, Stevens EM et al (1998) A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects? Am J Hum Genet 62:1044–1051. https://doi.org/10.1086/301825

Willnow TE, Hammes A, Eaton S (2007) Lipoproteins and their receptors in embryonic development: more than cholesterol clearance. Dev (Camb Engl) 134:3239–3249. https://doi.org/10.1242/dev.004408

Wu MC, Joubert BR, Kuan P et al (2014) A systematic assessment of normalization approaches for the Infinium 450 K methylation platform. Epigenetics 9:318–329. https://doi.org/10.4161/epi.27119