Dietary modification, penetrance, and the origins of congenital malformation

Molecular genetic diagnostics refers to the branch of clinical medicine dedicated to finding the proximate mutational cause of an abnormality (1). More for practical than theoretical reasons, it generally assumes that a genetic variant is both necessary and sufficient to cause the defect, and hence that there is no need to account for environmental conditions. An article by Cuny et al. (2) in PNAS challenges this view.

In syndromes where two or more phenotypes occur together, it is increasingly the case that a handful of known genes are responsible, so sequencing just those genes is sufficient to diagnose the cause. Even in idiopathic congenital abnormalities, for example neuromuscular diseases or craniofacial malformations, targeted panels of several hundred commonly mutated genes may suffice (3). In the next decade, it seems inescapable that whole-genome sequencing will become standard practice, and the conventional wisdom is that, once tens of thousands of cases of each condition have been characterized, we will know precisely which genes lead to which diagnoses.

There are, however, a few holes in this paradigm. A major one is that diagnostic yields, namely the proportion of cases solved by gene sequencing, are typically “only” in the range of 30 to 40% (4, 5). With 2 or 3% of people affected by a congenital malformation, this corresponds to thousands of diagnoses every year, and importantly overturns many misdiagnoses and thus changes treatment decisions even for unsolved cases. Another one is diagnostic uncertainty, since the definitions of pathogenic variant, likely pathogenic variant, and variant of unknown significance, depend on levels of evidence that remain in flux despite extensive clinical genetic guidelines (6, 7). We now know that there is extensive variability in how tolerant different genes are to mutations, including frameshifts or premature stop codons that may knock out the protein yet have little phenotypic impact. While databases of clinical variants continue to expand and improve (8), and bioinformatic filters designed to prioritize causal mutations become more sophisticated, in a large proportion of cases it remains unclear whether a specific genetic defect is necessary and sufficient to cause the abnormality.

A surprise that has emerged from the past several years of whole-genome analysis is that, except in the presence of consanguinity, the majority of congenital defects are due to dominant mutations rather than the inheritance of recessive alleles in homozygotes or compound heterozygotes (9). De novo variants can be haploinsufficient or gain of function, and if they are pathogenic then they are unlikely to be transmitted to future generations and hence only occur at very low frequencies. It seems reasonable to conclude that they are for the most part fully penetrant, that is, that they are only seen in cases. Monitoring presence or absence in affected or unaffected

---

**Fig. 1. Evidence of the potential for genotype-by-environment interaction (GxE) in congenital NAD deficiency.** (A) There are three pathways involved in synthesis of NAD pools. The de novo kynurenine pathway from tryptophan involves over a dozen enzymes, two of which catalyze enzymatic steps involving 3-hydroxyanthranilic acid, while a third is the final step. Biallelic mutations in all three of genes have been shown to cause VCRL syndrome. (B) Cuny et al. (2) tested various combinations of dietary supplementation of tryptophan in water at standard levels, reduced (600 mg/L) or low (500 mg/L) with heterozygosity for a null allele of the Haao gene. Bar graphs show the proportion of dead embryos (gray) or malformed embryos (dark orange, at least two defects; light orange, one defect) in each combination. Low tryptophan levels can cause phenocopies in wild-type mice, whereas reduced gene and nutrient dosage leads to GxE.
relatives is thus a useful criterion for establishing pathogenicity, but only so long as the effect of the allele is not condition specific.

There are three main reasons why this assumption of complete penetrance may be violated. The first is that genetic backgrounds are potent modifiers of allelic effects. We know this from model organisms where mutations can be introduced into different backgrounds—for example, even the dramatic Antennapedia mutation that grows legs in place of antennae in flies can be completely suppressed this way (10). More recently, a genome-wide association study of diverse congenital abnormalities has established a strong background genetic contribution to seemingly monogenic conditions (11). The second is stochasticity, the vagaries of development. It is not unusual for identical twins to be discordant for conditions from alopecia to autism, which although sometimes also attributed to epigenetics (12), is an often overlooked reminder that mutations need not be deterministic. The third is the environmental circumstances, which in humans mostly refers to toxin exposure or nutrition, which brings us to the paper in PNAS by Cuny et al. (2) at the Victor Chang Cardiac Research Institute in Sydney, NSW, Australia.

Cuny et al. (2) start with the observation they made a few years ago (13) that multiple cases of a syndrome characterized by a combination of vertebral, cardiac, renal, and limb defects (VCRL1 and VCRL2), now also known as congenital nicotinamide adenine dinucleotide (NAD) deficiency disorder (OMIM 617660 and 617661), can be attributed to biallelic loss of function of the genes HAAO and KYNU, which encode essential enzymes involved in the kynurenine pathway of NAD synthesis. More recently, NADSYN1 has been added to the list (14). They reasoned that if low levels of fetal NAD are responsible, then dietary restriction of the NAD precursors tryptophan and vitamin B3 ought also to promote these birth defects. Indeed, when they fed pregnant mice on a vitamin-depleted diet supplemented with 400 mg/L or less of tryptophan in the drinking water, fetal demise ensued, and increasing the levels to 500 mg/L restored some live births but with over half having multiple organ malformations. This is a classic case of the well-known phenomenon of a phenocopy, namely an environmental perturbation that mimics the phenotype of a known mutation.

Next, they further reasoned that the combination of a milder reduction of tryptophan with heterozygous loss of one of the genes involved in NAD biosynthesis might also generate the spectrum of congenital malformations. A careful set of analyses summarized in Fig. 1 show that this hypothesis is clearly validated. At a level of 600 mg/L of tryptophan, maternal heterozygosity at the Haa0 locus, a genetic constitution that is usually sufficient to support normal development, becomes sensitized to production of the syndrome. Furthermore, the birth defects correlate with measured levels of NAD in the maternal liver as well as whole embryos at embryonic day 11.5. Importantly, it is the maternal genotype that matters, likely because the fetal organs responsible for most of the embryonic NAD requirements, the liver and kidney, are not yet metabolically mature at the critical time for heart, vertebral, and limb organogenesis. This is then a classic case of genotype-by-environment interaction, where the penetrance of the genotype is conditional on the environmental, in this case maternal nutritional, status.

How common is this type of contribution to congenital malformation? The short answer is that, as far as we know, it is rare but it may well turn out to be quite common if researchers and clinicians look more systematically. Critics will note that this study does not actually document any cases of the human syndrome that have yet been attributed to a heterozygous mother experiencing NAD deficiency during pregnancy. On the other hand, some of the earliest human genetic investigations, those carried out by Sir Archibald Garrod (15), established that some inborn errors of metabolism can be corrected postnatally by dietary intervention, giving rise eventually to the Guthrie test for phenylketonuria, and these days for two dozen correctable metabolic insufficiencies evaluated routinely in all newborns. The conditions under which genotype-by-environment interactions in utero may cause birth defects have thus been known for decades. It is possible that the levels of NAD deficiency characterized in Cuny et al.’s experiments are so far below normal intake that these results are not relevant. However, the authors’ computations suggest that dietary tryptophan levels overlap the range in which a few percent of mothers may be at risk, particularly considering elevated metabolic requirements during pregnancy, and especially in diabetics and in the presence of inflammation (16).

Needless to say, these results do not imply that levels of recommended daily vitamin consumption need to be revisited. Nor do they suggest that a significant proportion of cases of congenital malformation that remain undiagnosed may actually be phenocopies. However, there are at least three implications of the research that are worth highlighting.

First, perhaps it is time to consider comprehensive screens for genotype-by-environment-induced birth defects in mice. This is in fact the second situation that the Dunwoodie group have uncovered. In 2012, they (17) showed that severe scoliosis can sometimes be attributed to the combination of fetal hypoxia and mutations in Notch pathway genes that regulate somite formation. Notably, fetal hypoxia secondary to bradycardia due to maternal intake of the antiarrhythmia drug dofetilide also interacts with Notch genotypes to cause congenital heart defects in mice (18). We do not yet have a compendium of human genotype-by-fetal environment interactions, but the discovery of potentially hundreds more potential combinations could have meaningful public health implications for isolated as well as syndromic birth defects.

Second, epidemiological scans are called for. Cuny et al. (2) report that there are 341 loss-of-function alleles present in the 17 known genes predicted to influence NAD biosynthesis in the gnomAD database of 141,456 exome and genome sequences of relatively healthy adults, along with another 3,000 or so missense alleles of unknown functional significance.

Cuny et al. report that there are 341 loss-of-function alleles present in the 17 known genes predicted to influence NAD biosynthesis in the gnomAD database of 141,456 exome and genome sequences of relatively healthy adults, along with another 3,000 or so missense alleles of unknown functional significance.
should be readily ascertainable. If an enrichment is found, it would suggest that close monitoring of nutrition in at-risk carrier mothers would be the type of personalized and predictive intervention that advocates of genomic health call for.

Third, this research points toward a time when the simplistic labeling of variants as pathogenic or likely pathogenic is replaced by databases that provide estimates of penetrance, expressivity, and modifiability at the genic, and eventually allelic, level (21). The current practice of using the instance of unaffected heterozygotes to exclude the possibility that a rare variant can be responsible for a condition is well justified in reducing false positives and imposing much needed rigorous clinical genetic standards. However, we are conceivably missing a large proportion of variants of unknown significance that have reduced penetrance (22), or which only act in combination with an unrecognized nutritional or other uterine deficiency. Even if that proportion of undiagnosed cases is just a few percent, there is no reason that modest steps cannot be taken to reduce the risk of congenital malformations for millions of expectant parents.

Acknowledgments
Research in G.G.’s laboratory is supported mainly by NIH grants from the National Human Genome Research Institute (R01-HG008146) and National Institute of Diabetes and Digestive and Kidney Diseases (R01-DK119991).