Indian J Med Res 134, November 2011, pp 588-590

Commentary

The hunt for genetic risk among Mysore, south India patent ductus arteriosus patients

Of the estimated 21 million babies annually born premature or with serious birth defects worldwide, more than 4 million die before the age of 5[1]. Although they constitute a diverse anatomical grouping, cardiac abnormalities taken together constitute the most common congenital malformations with frequencies around 1 per cent. Serious life threatening complications can occur from many cardiac structural defects, particularly those that involve an abnormal mixing of oxygenated with oxygen-depleted blood supplies. One such cardiac malformation is patent ductus arteriosus (PAD), a defect that arises from differences in the routing of foetal blood before and after birth. While still developing in the womb, blood circulation bypasses the foetus' lungs (which do not have access to their own oxygen supply) by connecting the pulmonary artery (normally supplying blood to the lungs) with the aorta (supplying blood to the body) through a foetal heart structure known as the ductus arteriosus. After birth, this blood vessel is no longer needed and closes off, since the infant's lungs now take on the role of supplying oxygen. Closure of this blood vessel ensures the normal separation of oxygenated and de-oxygenated blood flow after birth. In PAD, the ductus arteriosus fails to close following birth, and the normal segregation of blood flow is disrupted, often with serious consequences.

In the accompanying article in this issue[2] these authors investigated the role of five previously reported mutations in the gene encoding Transcription Factor AB 2 B (*TFAP2B*) in PAD patients from Mysore, south India. While their study population is not sizeable, and therefore, their results should be considered preliminary, these authors were not able to report the detection of any of the candidate gene alterations investigated in their study population. These results underscore the frustration often experienced by today's population geneticists due to the difficulties often encountered in identifying disease risk. A few remaining known *TFAP2B* mutations, previously associated with PAD but not investigated in this study, may still be identified in this population, but it is unlikely that all the cases would be due to them. Several other possibilities seem more probable. More than one gene is usually associated with a given diseased state. Defects in any one of several transcription factor genes, for example, are known to play a role in atrial septal defects, ventricular septal defects, tetrology of Fallot, and other cardiac malformations[3]. There may be a yet to be identified genetic predisposition to PAD which results in risk only when in combination with environmental factors such as vitamin deficiencies or toxic exposures *in utero*, as reported for several other birth defects[4-6]. *TFAP2B* may still play an important role in the occurrence of PAD among Mysore patients, but a different constellation of population specific mutations may be involved as is the case for myocilin mutations among 5 different study populations with glaucoma[7]. Lastly, copy number variations (CNVs) need to be considered. Haploinsufficiency resulting from hemizygosity is a common pathogenic mechanism among transcription factor genes[8], and may play an unreported role in the occurrence of PAD. Notably, the genotyping results for *TFAP2B* hemizygous individuals would be indistinguishable from those of homozygous normal genotypes by single nucleotide alteration approaches employed in this work. CNVs involving duplications must also be considered.

If limited to exons, Sanger sequencing of *TFAP2B* is affordable enough and could be conducted for this cohort to reveal the presence of any previously reported *TFAP2B* alterations not investigated in this study as well as the occurrence of novel *TFAP2B* mutations not yet reported by others. Additional technologies will be

needed if other known cardiovascular developmental genes or CNVs of *TFAP2* or other loci are to be considered. Two popular approaches commonly employed to evaluate CNVs involve either array comparative genomic hybridization (array CGH) analysis or multiplexed ligation probe amplification (MLPA). Array CGH is useful for detecting CNVs involving duplications/deletions of several hundred kb or larger, and when the need is to identify new candidate regions conferring risk. In this approach, a test sample DNA from an affected individual and a reference sample from a healthy control are labelled with complementary fluorescent dyes (usually green and red), co-hybridized to a DNA array containing probes for the entire human genome, and interrogated by laser scanning optics. The signal intensities are then aligned against the human genome map. A balance between the fluorescent signals of test and reference sample hybridizing to the array indicates the absence of CNVs. An overabundance of signal from the test sample indicates a duplication for the affected individual within the region of the genome under scrutiny, whereas an overabundance of signal from the reference DNA suggests that the test sample harbours a deletion. Since the sizes of deletions/duplications detected by this approach often run into the megabases, array CGH will merely identify a candidate region, which often may contain many promising candidate genes. Further investigations would then need to be conducted to determine which gene(s) within the candidate region can be associated with risk. MLPA on the other hand, is useful when the deletion/duplication regions are much smaller, as small as a single exon, and when specific known candidate genes are to be queried. In this approach pairs of hybridization probes are annealed in a contiguous fashion to sample DNA in a manner which satisfies the biochemical requirements of a ligation junction. Multiple non-overlapping ligation probe pairs, with electrophoretically distinguishable lengths and containing universal primer annealing tags for amplification, are designed for target exons within each candidate gene. Following incubation with a DNA ligase and PCR amplification with fluorescently labelled primers, the resulting amplified ligation products are examined by laser scanning electrophoresis and the integrated peak areas are compared with similarly treated control samples from healthy individuals to determine the presence of duplications or deletions among the cases. MLPA could easily be conducted to determine if any of Mysore patients are hemizygous for *TFAP2B*.

A fourth technology that could be used to further investigate genetic risk among these PDA patients is next generation DNA sequencing. Although still too expensive to routinely sequence complete genomes for entire cohorts of study subjects, large regions of the genome can be investigated for dozens to hundreds of individuals at reasonable costs. An attractive feature of next generation sequencing shared with Sanger sequencing is that both known and previously unreported alterations can be identified employing a single technology. However, an added advantage of next generation platforms is that these are more sensitive than the older Sanger chemistry, allowing the detection of rare sequence variants even within mixed samples[9], and of CNVs at high resolution[10]. In addition, next generation platforms are less expensive for the investigation of large candidate regions of the genome and, therefore, could be used to simultaneously investigate the role of all known cardiac developmental genes for simple nucleotide alterations, CNVs and other types of rearrangements.

A thorough epidemiological investigation is needed to determine the role of any potential gene-environment interactions. Detailed interviews with the mothers concerning perinatal maternal health that includes dietary information, vitamin supplementation, use of medications or possible exposures to toxic substances such as cigarette smoke in the months before and during early pregnancy must be conducted. Such studies however, can be both expensive and difficult to conduct since it requires additional contact with families that may have relocated since the time of the child's birth and therefore, may be challenging to complete. Similar data must be collected from mothers who gave birth to healthy babies for comparison. Once the epidemiological database has been constructed, and an association with risk can be determined for vitamin deficiencies, toxic exposures or other environmental factors, genetic data for candidate loci must be collected. Genes regulating vitamin absorption and metabolism as well as those involved in detoxification pathways are the most obvious candidates, but genes not encoding these activities have been evaluated with some success in other studies of gene-environment interactions[11-13] and, therefore, cardiac developmental genes could reasonably be included. Associations with risk for genes alone, environmental factors alone and gene-environment combinations may then be calculated.

Further investigation of this study population may ultimately reveal one of these mechanisms as

the explanation for PAD among this southern Indian population. For now, however, it is safe to say that the hunt has just begun.

David M. Iovannisci
Birth Defects Research Lab
Children's Hospital & Research
Centre Oakland
5700 Martin Luther King Jr. Way
Oakland, CA 94609, USA
Diovannisci@chori.org

References

1. The March of Dimes. Available from: *http://www.marchofdimes.com/mission/globalprograms.html*.
2. Kusuma L, Dinesh SM, Savitha MR, Krishnamurthy B, Narayanappa D, Ramachandra NB. Mutations of *TFAP2B* in congenital heart disease patients in Mysore, south India. *Indian J Med Res* 2011; *134* : 621-6.
3. Nemer M. Genetic insights into normal and abnormal heart development. *Cardiovasc Pathol* 2008; *17* : 48-54.
4. Carmichael SL, Shaw GM, Yang W, Iovannisci DM, Lammer E. Risk of limb deficiency defects associated with NAT1, NAT2, GSTT1, GSTM1, and NOS3 genetic variants, maternal smoking, and vitamin supplement intake. *Am J Med Genet A* 2006; *140* : 1915-22.
5. Lammer EJ, Shaw GM, Iovannisci DM, Van Waes J, Finnell RH. Maternal smoking and the risk of orofacial clefts: Susceptibility with NAT1 and NAT2 polymorphisms. *Epidemiology* 2004; *15* : 150-6.
6. Torfs CP, Christianson RE, Iovannisci DM, Shaw GM, Lammer EJ. Selected gene polymorphisms and their interaction with maternal smoking, as risk factors for gastroschisis. *Birth Defects Res A Clin Mol Teratol* 2006; *76* : 723-30.
7. Fingert JH, Heon E, Liebmann JM, Yamamoto T, Craig JE, Rait J, *et al*. Analysis of myocilin mutations in 1703 glaucoma patients from five different populations. *Hum Mol Genet* 1999; *8*: 899-905.
8. Seidman JG, Seidman C. Transcription factor haploinsufficiency: when half a loaf is not enough. *J Clin Invest* 2002; *109* : 451-5.
9. Thomas RK, Nickerson E, Simons JF, Janne PA, Tengs T, Yuza Y, *et al*. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med* 2006; *12* : 852-5.
10. Chiang DY, Getz G, Jaffe DB, O'Kelly MJ, Zhao X, Carter SL, *et al*. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat Methods* 2009; *6*: 99-103.
11. Shaw GM, Iovannisci DM, Yang W, Finnell RH, Carmichael SL, Cheng S, *et al*. Endothelial nitric oxide synthase (NOS3) genetic variants, maternal smoking, vitamin use, and risk of human orofacial clefts. *Am J Epidemiol* 2005; *162* : 1207-14.
12. Wu T, Liang KY, Hetmanski JB, Ruczinski I, Fallin MD, Ingersoll RG, *et al*. Evidence of gene-environment interaction for the IRF6 gene and maternal multivitamin supplementation in controlling the risk of cleft lip with/without cleft palate. *Hum Genet* 2010; *128* : 401-10.
13. Jianyan L, Zeqiang G, Yongjuan C, Kaihong D, Bing D, Rongsheng L. Analysis of interactions between genetic variants of BMP4 and environmental factors with nonsyndromic cleft lip with or without cleft palate susceptibility. *Int J Oral Maxillofac Surg* 2010; *39* : 50-6.