Whole exome sequencing to uncover genetic variants underlying congenital cystic adenomatoid malformations

PK Tam*

KEY MESSAGES

1. Congenital pulmonary airway malformation is a risk factor for paediatric adenocarcinoma of the lungs.
2. Both point mutations in coding sequences and copy number variants involving gene deletions and duplications are recurrently mutated in patients with congenital pulmonary airway malformation but not in the general population.
3. Congenital pulmonary airway malformation is genetically heterogeneous with different mutated genes in different patients. Mutations in more than one gene and/or copy number variants involving several genes are required to give rise to the developmental anomaly.
4. Despite the diversity, mutated genes encode interacting proteins that are members of the same cancer pathway may be a potential therapeutic target.

Hong Kong Med J 2019;25(Suppl 5):S36-8
HMRF project number: 01121576

PK Tam

Department of Surgery, The University of Hong Kong

* Principal applicant and corresponding author: paultam@hku.hk

Introduction

Congenital pulmonary airway malformation (CPAM) is characterised by fetal hamartomatous pulmonary lesions that result from abnormal overgrowth of tracheal, bronchial, bronchiolar, or alveolar tissues. Most CPAMs present antenatally. The clinical presentation ranges from the most serious phenotype that may result in hydrops fetalis and termination of pregnancy to mild phenotype with possible resolution. Affected new-borns present with severe respiratory distress or remain asymptomatic until later in life. Surgical resection is the definitive treatment.[1] Primary adenocarcinomas of the lung in paediatric patients are preponderantly found in conjunction with congenital CPAM and are considered a risk factor for paediatric cancer.[2]

CPAM is caused by a defective branching morphogenesis of the lung at different developmental stages; however, the trigger for this developmental defect is unknown.[1] Data on the molecular basis underlying CPAM are scant and consist of gene expression analyses in fetal or postnatal resected human CPAM tissues or in animal models. Yet, these studies have identified deregulation of genes/proteins crucial for lung morphogenesis and patterning. Thus, alteration of the signalling pathways controlling lung development is likely the mechanism underlying CPAM. Plausibly, DNA alterations in gene members of the involved pathways could lead to CPAM. The disorder has not been linked to race, maternal age, or exposures.[1]

There are gaps in knowledge and uncertainties/controversies in management of CPAM. No genetic study has been conducted on CPAM, likely because there is no evidence for a classic genetic inheritance of the disorder together with the scarcity of patients. Nonetheless, identical twins affected with CPAM born to unaffected parents have been reported.[3] CPAM appears mainly sporadically at a very low incidence (1/8300 and 1/35,000 live births). Thus, we aim to investigate the genetics underlying CPAM under the hypothesis that rare *de novo* or recessive inherited damaging genetic variants in genes governing the development of airways may trigger the disorder and account for the sporadic presentation and scarcity of CPAM.

Results

Sample processing

All patients had a normal karyotype. One trio was excluded owing to an accidental sample duplication. One patient who had bilateral lung hypoplasia was excluded because it might suggest different disease aetiology. Therefore, 18 trios were analysed.

De novo exonic variants

We identified 13 non-synonymous novel *de novo* damaging variants in 13 genes distributed among 11 patients. We then queried Clin Var, Online Mendelian Inheritance in Man database, COSMIC, and Mouse Genome Informatics database to assess the involvement of CPAM genes carrying those de novo mutations in other human disorders. There were 23 rare (MAF, ≤0.5%) or *de novo* coding DNA sequence mutations in 23 genes. Eight of the 23 genes were involved in human genetic diseases, and a novel protein truncating variant was detected in CEP295

(centrosomal protein 295). In addition, 21 of the 23 genes were involved in either adenocarcinoma or small cell carcinoma. *CACNA1H*, *EPG5*, *TRIP12* have been reported to be >50 times in adenocarcinoma samples in COSMIC database. As to the phenotype caused by these mutated genes in mouse/human orthology with phenotype annotations, 16 of 23 genes have been reported in different developmental phenotypes. CPAM patients with mutations in genes known to be involved in other congenital human disorders were clinically re-assessed, and no other conjunctional congenital human disorders was observed.

Inherited exonic variants: homozygosis, compound heterozygosity and di-genic models

We then assessed patients for inherited damaging variants acting in a recessive manner including homozygous, compound heterozygous, and digenic inheritance mutations. For rare homozygous mutations, we detected 24 mutations in 23 genes. Among 24 homozygous mutations, 20 were predicted to be deleterious by logistic regression in 11 CPAM patients. After assessing the involvement of the CPAM genes carrying rare homozygous mutations in other human disorders, two variants were reported in ClinVar including the homozygous mutations on *LHB* and *PGAM2* genes. Also, 8 of 23 genes were reported in Online Mendelian Inheritance in Man database, and 12 of 23 genes were reported in Mouse Genome Informatics database. All of the rare homozygous mutated genes were involved in adenocarcinoma according to the COSMIC database. Importantly, patient CCAM11C and CCAM22C had identical rare homozygous damaging mutation in *LTBP2* gene. According to ProteomicsDB, *LTBP2* has highest RNA expression level in lung among other organs. For compound heterozygotes (CH), the events with at least one deleterious allele were identified in 15 sharing/recurrent CH genes and distributed among 16 CPAM families (Table). As above, 13 of 15 CH genes were reported in adenocarcinoma in COSMIC dataset. Similarly, we identified 20 recurrent, physical protein-protein interaction pairs with digenic inheritance mutations distributed among 16 families. One protein inherited at least one rare non-synonymous variants from the father, and the other protein has another inherited rare non-synonymous mutation from the mother. Only recurrent interactive pairs were selected.

A gene with different damaging inheritance variants in different individuals (recurrently mutated gene) was identified pathogenicity. In total,

TABLE. Genetic profile of the rare damaging genetic variations in patients with congenital pulmonary airway malformation (CPAM)*

Trio	*De novo*	Homo	CH	Digenic	Recurrent genic copy number variant*
CPAM11		***LTBP2,*** *PGAM2*	*CCDC150, CUBN,* ***TTN***	*TG, LRP2; FASN, AOAH;* ***KIAA0232,*** *OBSL1*	
CPAM13	*MICAL2, SCYL1*	***ASCC1***	*DNAH17, PNPLA7,* ***TTN***	*LASP1,* ***PRKG1; C5, C7;*** *HIST1H2AK, PCF11; SEMA4A, PLXNB1; BEAN1, NEDD4; PRKDC, SMG1*	***C5***
CPAM14				*LASP1,* ***PRKG1;*** *TNK1, SRRM2*	
CPAM15		*AR, BNC1, KIR2DS4, MYL10*	***CEP128, CEP295,*** *CUBN, FOCAD*	*CFD, HPX; MLH1, MSH3; FASN,* ***TTN; OBSCN, TTN;*** ***KIAA0232,*** *OBSL1*	***CEP128, FAT3, PRKG1***
CPAM16	*CHPF, CAMKK2*	*GLI1*	*PAPLN, SPTBN5,* ***TTN***	*MAP2, KIF21B*	
CPAM18	*IMPDH1*	*PHKA1*		*FASN,* ***TTN;*** *FASN, AOAH; FN1, AOAH;* ***OBSCN, TTN***	
CPAM19	*CASQ2*	*ZNF467*	***TTN***	***OBSCN, TTN;*** *FASN, AOAH; FN1, AOAH*	
CPAM2	*TTC27*	*LCN1*	*DNAH17, PNPLA7*		
CPAM20			*AHNAK2,* ***FAT3,*** *MYBBP1A*	*MAP2, KIF21B;* ***C5, C7;*** *CFD, HPX; SEMA4A, PLXNB1*	
CPAM22	*EPG5*	***LTBP2,*** *TBCC*	***CEP128, FAT3,*** *SPTBN5,* ***SYNE1***	*MLH1, MSH3*	
CPAM23	***CEP295***		*OBSCN, TTN*	*TLN2,* ***KIAA0232***	***C5***
CPAM24	*FPGS*		*MYBBP1A*	*HIST1H2AK, PCF11*	
CPAM25	*CKB*	*UGT1A3*	***C7, OBSCN, TTN***	*TG, LRP2; CARD6,* ***SYNE1;*** *TLN2,* ***KIAA0232***	
CPAM28	*TRIP12*	*GPR84, RFX6*	***CEP295***	*TNK1, SRRM2*	
CPAM30		*EDA2R, EGFL6, FXYD4, GANAB, PEPD*	*AHNAK2, CCDC150,* ***TTN***	*BTRC, IKBKB; CARD6,* ***SYNE1;*** *PRKDC, IKBKB*	
CPAM31			*PAPLN,* ***SYNE1, TTN***		
CPAM6	*ING1*	*HEPH, LHB, SMAD7*	*DNAH17, FOCAD*	*BEAN1, NEDD4*	***ASCC1***
CPAM9			***C7, TTN***	*PRKDC, SMG1; PRKDC, IKBKB*	

* Genes contained in copy number variants. **Bold** indicates that the genes are recurrently mutated in different inheritance modes.

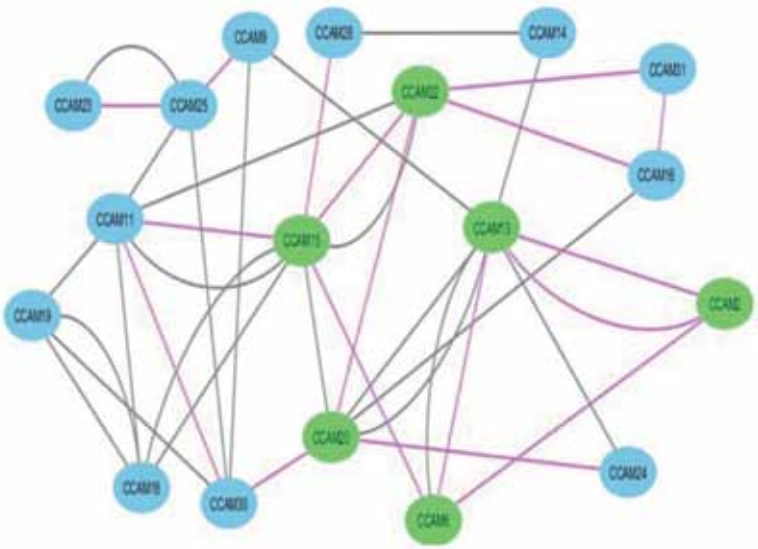

FIG. Shared genetic components among patients: green nodes represent patients sharing more than 3 CH variants. Gray connectors represent patients connected share variants in interacting proteins. Purple connectors represent patients connected share CHs.

we identified 81 genes with rare mutations including recurrently mutated gene in either CH or digenic inheritance modes, and as well as *de novo* or in homozygosis. Importantly, the alleles of recurrently mutated genes are not necessarily the same (gene mutated at different sites).

Gene-based and gene-set-based association test

To boost power for association analysis, we attempted to account for biological or functional relatedness by resorting to pathway (MsigDB) databases to group rare variants for association test. Gene-set-based (pathways) association tests revealed only one gene set reaching marginal significance in burden test (P<9.04391E-05). The pathway corresponded to that of medulloblastoma tumours in mice. The top five pathways included genes with unmethylated DNA in lung cancer samples (38 genes). The case-control association test indicated the same deleterious direction between CPAM and control samples with P values of $3.57e^{-3}$, $1.96e^{-3}$, $2.71e^{-3}$ in SKAT, SKATO, and Burden tests, respectively. All CPAM patients were 'genetically' linked (Fig).

Searching functional overlap among genes

To understand how these mutated genes in our patients contributed to CPAM and to consider whether our findings fit into any pathological process, we performed gene/pathway-set enrichment analyses and carefully examined the genetic profile or mutational load of each patient. The gene-set enrichment test indicates several cancer-related gene-sets including Ewing's sarcoma and prostate cancer ($P=1.55e^{-6}$ and $P=2.28e^{-6}$, respectively). The genetic profile of each of the 18 CPAM patients is shown in the Table.

Discussion

We investigated the lesions in the genome that might be underlying CPAM. We used a straight forward design according to the presentation and incidence of the disease in the paediatric population by studying both coding sequence mutations and copy number variants. The highlights of our findings are: (1) most mutated genes are implicated in cancers of the lung and congenital soft tissue sarcomas; (2) *ASCC1* is highly expressed in lung cancer and is implicated in oesophageal adenocarcinoma and is found recurrently mutated twice, with one event being a homozygous point mutation and the other being a complete gene deletion; (3) mutations in genes encoding interacting proteins member of pathways implicated in cancer development were also recurrently detected; and (4) mutation recurrence was observed for both point coding DNA sequence mutations and copy number variants.

Despite heterogeneity, our findings provide a pathway to look for a therapeutic target. Although our data need to be replicated, and functional assays are needed, it is a lead that hopefully the rest of the scientific community will follow.

An inherent limitation to this project (and to any other project involving a genetic study of a rare genetic heterogenic and oligo/polygenic disorder) is the small sample size and/or the lack of large pedigrees where the disease segregated. When different genes are thought to underlie the same disorder, a large sample size is needed so that genes recurrently mutated can be identified. Intuitively, the different genes underlying the disorder should be connected through a pathway, yet, not all pathways or connectivity are known. Replication has not been done as the number of samples is insufficient.

Acknowledgements

This study was supported by the Health and Medical Research Fund, Food and Health Bureau, Hong Kong SAR Government (#01121576). We thank the patients who participated in the study.

References

1. Wilson RD, Hedrick HL, Liechty KW, et al. Cystic adenomatoid malformation of the lung: review of genetics, prenatal diagnosis, and in utero treatment. Am J Med Genet A 2006;140:151-5.
2. Kayton ML, He M, Zakowski MF, et al. Primary lung adenocarcinomas in children and adolescents treated for pediatric malignancies. J Thorac Oncol 2010;5:1764-71.
3. DeBoer EM, Keene S, Winkler AM, Shehata BM. Identical twins with lethal congenital pulmonary airway malformation type 0 (acinar dysplasia): further evidence of familial tendency. Fetal Pediatr Pathol 2012;31:217-24.