

Genome-wide copy number variation study in anorectal malformations

Emily H.M. Wong^{1,†}, Long Cui^{2,†}, Chun-Laam Ng^{2,†}, Clara S.M. Tang^{1,2}, Xue-Lai Liu², Man-Ting So², Benjamin Hon-Kei Yip^{1,2}, Guo Cheng², Ruizhong Zhang², Wai-Kiu Tang², Wanling Yang⁵, Yu-Lung Lau⁵, Larry Baum⁷, Patrick Kwan⁷, Liang-Dan Sun^{8,9}, Xian-Bo Zuo^{8,9}, Yun-Qing Ren^{8,9}, Xian-Yong Yin^{8,9}, Xiao-Ping Miao¹⁰, Jianjun Liu¹¹, Vincent Chi-Hang Lui^{2,4}, Elly Sau-Wai Ngan^{2,4}, Zhen-Wei Yuan¹², Shi-Wei Zhang¹³, Jinglong Xia¹³, Hualong Wang¹⁴, Xiao-bing Sun¹⁵, Ruoyi Wang¹⁵, Tao Chang¹⁶, Ivy Hau-Yee Chan², Patrick Ho-Yu Chung², Xue-Jun Zhang^{8,9}, Kenneth Kak-Yuen Wong², Stacey S. Cherny^{1,6}, Pak-Chung Sham^{1,3,4,6}, Paul Kwong-Hang Tam^{1,2,4} and Maria-Mercè Garcia-Barcelo^{24,*}

¹Department of Psychiatry, ²Department of Surgery, ³Center for Genomic Sciences, ⁴Centre for Reproduction, Development and Growth, ⁵Department of Paediatrics and Adolescent Medicine of the Li Ka Shing Faculty of Medicine and ⁶State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong SAR, China ⁷School of Pharmacy, the Chinese University of Hong Kong, Hong Kong SAR, China, ⁸Institute of Dermatology and Department of Dermatology, No.1 Hospital, Anhui Medical University, Anhui, China, ⁹State Key Laboratory Incubation Base of Dermatology, Ministry of National Science and Technology, Hefei, Anhui, China, ¹⁰Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, ¹¹Department of Human Genetics, Genome Institute of Singapore, Singapore, ¹²Department of Paediatric Surgery, Shengjing Hospital, China Medical University, Shenyang, China, ¹³Harbin Children's Hospital, Harbin, China, ¹⁴Changchun Children Hospital, Changchun, China, ¹⁵Department of Paediatric Surgery, Shandong Medical University, Shandong, China and ¹⁶Zhejiang Children's Hospital, Zhejiang, China

Received May 21, 2012; Revised October 3, 2012; Accepted October 19, 2012

Anorectal malformations (ARMs, congenital obstruction of the anal opening) are among the most common birth defects requiring surgical treatment (2–5/10 000 live-births) and carry significant chronic morbidity. ARMs present either as isolated or as part of the phenotypic spectrum of some chromosomal abnormalities or monogenic syndromes. The etiology is unknown. To assess the genetic contribution to ARMs, we investigated single-nucleotide polymorphisms and copy number variations (CNVs) at genome-wide scale. A total of 363 Han Chinese sporadic ARM patients and 4006 Han Chinese controls were included. Overall, we detected a 1.3-fold significant excess of rare CNVs in patients. Stratification of patients by presence/absence of other congenital anomalies showed that while syndromic ARM patients carried significantly longer rare duplications than controls ($P = 0.049$), non-syndromic patients were enriched with both rare deletions and duplications when compared with controls ($P = 0.00031$). Twelve chromosomal aberrations and 114 rare CNVs were observed in patients but not in 868 controls nor 11 943 healthy individuals from the Database of Genomic Variants. Importantly, these aberrations were observed in isolated ARM patients. Gene-based analysis revealed 79 genes interfered by CNVs in patients only. In particular, we identified a *de novo* *DKK4* duplication. *DKK4* is a member of the *WNT* signaling pathway which is involved in the development of the anorectal region. In mice, *Wnt* disruption results in ARMs. Our data suggest a role for rare CNVs not

*To whom correspondence should be addressed at: Department of Surgery, Li Ka Shing Faculty of Medicine of the University of Hong Kong, 1/F The Hong Kong Jockey Club Building for Interdisciplinary Research, 5 Sassoon Road, Pokfulam, Hong Kong, Hong Kong. Tel: +852 28315073; Fax: +852 28199621; Email: mmgarcia@hku.hk

[†]These authors contributed equally to the work.

only in syndromic but also in isolated ARM patients and provide a list of plausible candidate genes for the disorder.

INTRODUCTION

Anorectal malformations (ARMs, congenital obstruction of the anal opening) are among the most common birth defects requiring surgical treatment (2–5/10 000 live-births) (1) and carry a significant chronic morbidity. The condition is attributed to a defect in the proliferation of the embryonic rudiments that will form the distal end of the gut, and it is probably due to disorders in the expression of pattern determining genes. The spectrum of ARMs ranges from anal stenosis to imperforated anus with/without anal fistula to persistent cloaca, in which the intestinal and genitourinary tracts remain a common channel. ARMs might appear as part of the phenotypic spectrum of many chromosomal abnormalities (2–4) or monogenic syndromes (5,6).

The etiology of ARMs remains unknown. While environmental factors are not to be dismissed, several lines of evidence indicate that there is a genetic component (7). Indeed, even though ARMs appear mostly sporadically (no affected relatives), they also segregate within families with patterns of inheritance ranging from autosomal-dominant, X-linked, to autosomal-recessive (8–10). Moreover, higher risk of anal atresia/stenosis has been associated with consanguinity (11).

The approach currently being taken towards the discovery of genes involved in ARMs in humans is that of the analysis of candidate genes selected according to the data provided by (i) their role in syndromes that include ARMs as part of their spectrum; (ii) mutant mice/rat studies (12,13) as in most cases, mutations in the human orthologs give rise to similar or related phenotypes. However, while mice mutant for *Shh*, *Gli2* or *Gli3* display different congenital defects that include ARMs as a common feature, point mutations in the human orthologs (*SHH*, *GLI3*) are associated with syndromes or genetically heterogeneous disorders in which the ARMs phenotype is not always the norm. Failure to identify human genes underlying ARMs so far may be attributed to the reason that there is no single major gene, suggesting genetic and phenotypic heterogeneity.

To explore the genetic contribution to the pathogenesis of this condition, we investigated at genome-wide scale, single-nucleotide polymorphisms (SNPs) and copy number variations (CNVs) on Han Chinese ARM patients. Patients were also stratified into those presenting ARMs as an isolated feature and those presenting ARMs together with associated anomalies and analyzed accordingly.

RESULTS

Common variants do not contribute to the isolated ARMs phenotype

Common SNPs

After stringent quality control (QC) on genome-wide association study (GWAS) genotypes, 489 243 SNPs (average call rate 99.85%) that were successfully genotyped in 175 ARMs

cases and 2971 controls were tested for association to ARMs under additive, dominant and recessive models. After correcting for the first 12 principal components (i.e. correcting for population stratification), 48 SNPs reached association P -values $< 10^{-4}$ under the additive model (see Supplementary Material, Table S3). No SNP reached genome-wide significance ($P < 10^{-8}$). QQ-plot and Manhattan plot of the association test results were shown in Supplementary Material, Figures S2 and S3, respectively. The genomic inflation factor after principal components analysis (PCA) correction did not deviate from 1 ($\lambda = 1.0093$) indicating that chances of spurious associations due to the population substructure were minimal.

As PCA had revealed that two subpopulations of Han Chinese (Northern and Southern) existed in our data set (see Supplementary Material, Fig. S1), we then performed association tests separately to assess if susceptibility loci differed between the two subpopulations. When the northern Chinese patients ($n = 103$) were tested for ARMs association against northern Chinese controls ($n = 1411$), we identified 44 SNPs with association P -values of $< 10^{-4}$ under the additive model. As for the southern Chinese (patients: $n = 72$; controls: $n = 1560$) we identified 42 SNPs with P -values $< 10^{-4}$. Similar results were obtained when patients were stratified according to the presence/absence of additional anomalies or syndromes (data not shown).

We then proceed to replicate the signals detected in the discovery phase. After QC, 110 SNPs were successfully genotyped on 167 patients affected with isolated ARM (northern: $n = 81$, southern: $n = 86$) and 174 normal controls (northern: $n = 88$, southern: $n = 86$). Meta-analysis of all Han Chinese revealed that none of the combined P -values for the SNP tested reached genome-wide significant level.

Copy number polymorphisms

Global burden of CNPs were examined in all cases and controls in terms of length, frequency and number of genic regions overlapped. None of the global burden test results was statistically significant (see Supplementary Material, Table S6), thus suggesting that there is no enrichment of common CNVs in cases compared with controls. Similar results were obtained when the analysis was performed on stratified patients (isolated or syndromic patients with ARMs).

Lack of ARMs association with common susceptibility loci may also imply that the disorder results from variants whose frequency in the population is $< 1\%$ (rare variants). Using the same data set, we proceeded to study the contribution of rare CNVs to the ARMs phenotype as described in the following sections.

Rare CNVs contribute to ARMs

Global burden analysis: ARM patients are enriched with rare CNVs

Rare CNVs are known to play a more significant role in disease susceptibility than CNPs (common CNVs) (14,15).

Table 1. Global burden of RARE CNVs in ARMs cases and controls

	Rare CNVs with length <100 kb		Empirical <i>P</i> -values	Rare CNVs with length >100 kb		Empirical <i>P</i> -values
	Cases (<i>n</i> = 170)	Controls (<i>n</i> = 851)		Cases (<i>n</i> = 170)	Controls (<i>n</i> = 851)	
Total number of segments						
Deletion	538	2165		66	216	
Duplication	218	888		126	374	
Number of rare CNVs per sample						
Deletion	3.165	2.544	0.03945	0.3882	0.2538	0.021955
Duplication	1.282	1.043	0.007344	0.7412	0.4395	0.002307
Proportion of samples with one or more rare CNVs						
Deletion	0.9059	0.8261	0.00484	0.2647	0.2162	0.100801
Duplication	0.6941	0.6381	0.093875	0.4294	0.3467	0.025701
Total length of rare CNVs spanned per sample (in kb)						
Deletion	92.21	84.2	0.2479	314.6	268	0.234714
Duplication	69.67	57.38	0.01022	505.2	373.9	0.060274
Number of genic regions spanned by rare CNVs per sample						
Deletion	1.247	1.905	0.9369	0.4882	0.3384	0.15943
Duplication	0.8588	0.6616	0.026837	1.176	0.8461	0.063724
Number of genic CNVs per sample						
Deletion	0.5941	0.5347	0.09023	0.1118	0.1234	0.704262
Duplication	0.4529	0.3643	0.018863	0.3118	0.2526	0.068167

Statistical significance was inferred using permutation with 1 000 000 iterations. Rare CNVs are defined as CNVs that are observed in <1% samples of the data set. ARM cases (isolated: *n* = 126; syndromic: *n* = 44) are enriched with rare deletions and rare duplications by 1.3-fold each. Separate global burden tests were also performed by stratifying the isolated and syndromic ARM patients. Details are presented in the Supplementary Material, Table S7. Statistically significant values shown in bold.

Thus, global burden of rare CNVs (defined as CNVs that are observed in <1% samples of the data set) was compared between cases and controls using permutation tests (1 000 000 iterations). Patients were enriched with rare deletions and rare duplications by 1.3-fold each. Rare CNVs were classified into two groups in terms of size (length <100 kb or length >100 kb) and their distributions in patients and controls are tabulated in Table 1. Globally, more duplications were identified in ARM patients (CNV with length <100 kb: empirical *P*-value = 0.007344; CNV with length >100 kb: empirical *P*-value = 0.002307) when compared with controls.

Significant enrichment of rare duplications (CNV with length <100 kb: empirical *P*-value = 0.0187; CNV with length >100 kb: empirical *P*-value = 0.0064) and deletions (CNV with length <100 kb: empirical *P*-value = 0.0102; CNV with length >100 kb: empirical *P*-value = 0.0181) were observed in non-syndromic patients (isolated ARMs, *n* = 126), while interestingly, syndromic patients (*n* = 44, among which 15 have Down syndrome) were only modestly enriched with long duplications (length >100 kb, empirical *P*-value = 0.0490). The results were tabulated in Supplementary Material, Table S7a and b. Although the association tests are significant, it would be important to replicate the excess in deletions and duplications in an independent group of patients and controls of all ancestries available.

Chromosomal aberrations

Within the set of long CNVs, we examined closely those CNVs longer than 1 Mb which are referred to as chromosomal aberrations. Global burden analysis revealed that, overall, ARM patients have 3-fold more chromosomal aberrations (defined as longer than 1 Mb) than controls (*P* = 0.0368 for deletions and *P* = 0.00614 for duplications) even after excluding patients with Down syndrome (around 9% of our ARM

patients) (see Supplementary Material, Table S8a and b). The chromosomal aberrations in ARM patients also spanned more genic regions than those in controls. This applied to either deletions (empirical *P*-value = 0.00038) or duplications (empirical *P*-value = 0.000226).

Importantly, we identified 12 chromosomal aberrations (besides trisomy 21) that were unique to ARM patients as they were not identified in controls or in the normal individuals of the Database of Genomic Variants (DGV) (Table 2). Seven chromosomal aberrations encompassed genes that were not disrupted in controls. The chromosomal aberrations observed were checked against the Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER) which is a database of sub-microscopic chromosomal imbalances and clinical information for 6169 patients with developmental disorders. There were five aberrations that had also been found in DECIPHER patients with similar or ARM-related symptoms.

Interestingly, a 2.5 Mb heterozygous deletion on chromosome 22q11.21 was identified in one affected female (MG-IA162C, isolated ARM). Deletions involving 22q11.21 had been reported in a patient with VACTERL syndrome, which includes ARMs as part of the spectrum (16), and in five other syndromic ARM patients (17–19). Chromosomal aberrations on chromosome 22q11.21 were also reported in five DECIPHER patients with hindgut problems (Table 2), including one patient with isolated ARM and one syndromic patient with sacrum and kidney anomalies. Thus, it would appear that chromosomal aberrations involving this region are not only involved in syndromes with ARMs as part of the phenotypic spectrum but also in the isolated ARMs phenotype (20). CNVs that overlap known critical regions are likely to be pathogenic in nature. Importantly, these chromosomal aberrations were identified in patients with the ARM-isolated phenotype.

Table 2. Chromosomal aberrations that were only observed in ARM patients

Chr.	Chromosomal aberrations	Patient ID	Isolated (I) or syndromic (S)	Genes uniquely disrupted in cases	Patients with related symptoms listed in DECIPHER (patient record, type of CNV, phenotype)
1	1.3 Mb Duplication in 1q21.1	MG-IA201C	I	<i>ACP6</i>	record 983: DUP, absent uterus, fused labia, vaginal atresia
1	2.3 Mb deletion in 1q42.3–q43	MG-IA195C	I	<i>RBM34, MTR</i>	
1	8.1 Mb deletion in 1q43–q44	MG-IA87C	I	<i>FAM152A</i>	record 249405 ^b : DUP and DEL, megacolon/Hirschsprung syndrome, general abnormalities in heart
2	1.5 Mb deletion in 2q37.3	MG-IA33C	I		
5	3.7 Mb deletion in 5p15.33	MG-IA41C	I	<i>ISL1</i>	record 4119 ^b : DUP, megacolon/Hirschsprung syndrome, general abnormalities in heart
5	1.1 Mb duplication in 5q11.2	MG-IA147C	I ^a		record 1946: DEL, intestinal malrotation
7	1.4 Mb duplication in 7p21.3–p21.2	MG-IA370C	I		
7	6.7 Mb duplication in 7p11.1–7p11.21	MG-IA360C	I		
8	13.2 Mb duplication in 8p12–q11.21	MG-IA147C	I ^a	<i>ZNF703, ERLIN2, PROSC, BRF2, RAB11FIP1, GOT1L1, ADRB3, ASH2L, STAR, DDHD2, PPAPDC1B, FGFR1, TACC1, C8orf4, GOLGA7, GINS4, AGPAT6, NKX6-3, AP3M2, DKK4, C8orf40, CHRNA6, THAP1, CEBPD, SNAI2</i>	
11	8.4 Mb duplication in 11q14.3–11q22.1	MG-IA152C	I	<i>JOSD3, C11orf54, MED17, LOC390243, GPR83, PIWIL4, AMOTL1, CWC15, JMJD2D, SFRS2B, ENDOD1, FAM76B, MTMR2, CCDC82, JRKL</i>	
15	1.6 Mb duplication in 15q13.2–q13.3	MG-IA230C	I		
22	2.6 Mb deletion in 22q11.21	MG-IA162C	I	<i>SERPIND1, SNAP29</i>	record 252033 ^b : DUP, anal atresia/stenosis, general abnormalities in sacrum and kidneys; record 2366 ^b : DEL, anal atresia/stenosis, general heart abnormalities; record 249397 ^b : DEL, megacolon/Hirschsprung; record 622 ^b : DEL, absent uterus, general abnormalities in kidneys; record 1645 ^b : DEL, malformed uterus, renal agenesis, fusion of vertebrae

Chromosomal aberrations are defined as CNVs with length longer than 1 Mb. The patient IDs are listed to show the number of patients harboring the chromosomal aberrations. The chromosomal aberrations were also checked against the DECIPHER database for any other patients with similar or ARM-related symptoms.

^aPatient was diagnosed with autism at the age of 6.

^bThis is the only chromosomal aberration observed in this DECIPHER patient.

Rare CNVs that are statistically associated to ARMs

Following the global rare CNVs enrichment observed in the patients, rare CNV regions (CNVR; genic: $n = 457$ and non-genic: $n = 1982$) were defined (see Supplementary Material, Fig. S5) and individually evaluated for their associations with ARMs by using permutation tests. The number of CNVs within each defined CNVR was compared between patients and controls. We also performed the more appropriate for low counts Fisher's exact test, and included a Bonferroni correction for 2439 total rare CNV regions.

Using permutation tests, we identified three non-genic CNV regions that were statistically associated with ARMs on 7p14.1 (corrected empirical P -value = 0.000147; P -value

from Fisher's exact test after Bonferroni correction = 0.00779), 14q11.2 (corrected empirical P -value = 0.00137; P -value from Fisher's exact test after Bonferroni correction = 0.0482) and 1p34.2 (corrected empirical P -value = 0.0229; P -value from Fisher's exact test after Bonferroni correction = 0.594) (Table 3). All were hemizygous deletions.

Deletions on 7p14.1 (a 45 kb region) were observed in seven ARM patients (six isolated; one with bifid scrotum) but in none of the controls. This region [5.3 kb upstream of *TARP* (TCR gamma alternate reading frame protein)] overlaps a 411 bp CpG island and a transcription factor binding site (see Supplementary Material, Fig. S8A). Deletions on 14q11.2 (a 73 kb region) were observed in five ARMs cases

Table 3. Results of the top three rare CNV regions from the permutation test and Fisher's exact test

Chr.	Starting position (in hg18)	Ending position (in hg18)	Number of cases	Number of controls	Permutation test		Fisher's exact test	
					Empirical <i>P</i> -values	Empirical <i>P</i> -values corrected for all tests	<i>P</i> -values	<i>P</i> -values after Bonferroni correction
7	38285115	38330273	7	0	2.00×10^{-06}	0.000147	3.20×10^{-06}	0.00779436
14	21937715	22009307	6	0	0.000134	0.001372	1.98×10^{-05}	0.048239486
1	40794563	40804646	7	3	0.000240	0.022894	0.000243573	0.594073571

Genic and non-genic regions are defined as those that harbor at least one CNV in cases or controls. These three regions do not harbor any genes, i.e. non-genic regions. Permutation tests (1 000 000 iterations) were performed on 2439 rare CNV regions (1982 genic regions and 457 non-genic regions) by using PLINK. According to PLINK, the empirical *P*-values corrected for all the tests were calculated by comparing each observed test statistic against the maximum of all permuted statistics (i.e. over all regions) for each single replicate. We also performed the Fisher's exact test on these regions, and included a Bonferroni correction for 2439 total CNV regions.

(three isolated; one with bifid scrotum, one with Down syndrome) but not in our controls (see Supplementary Material, Fig. S8B). The 1p34.2 ARMs-associated deletion was detected in seven ARM patients (six isolated; one with heart and kidney anomalies) and in three control individuals (see Supplementary Material, Fig. S8C).

These three non-genic regions overlapped with CNVs listed in the DGV. Yet, given their associations with the ARMs phenotype, it is tempting to speculate that they affect regulatory sites and that may contribute to disease in conjunction with additional altered loci.

Rare CNVs unique to ARM patients

As rare CNVs are more likely to be pathogenic if they involve gene-rich regions and are only found in affected individuals, we proceeded with the identification of rare CNVs that were exclusive to ARM patients (see Supplementary Material, Fig. S6). This yielded 433 CNVs of which 342 were observed only once (non-recurrent) and 91 CNVs (distributed in a total of 35 CNVR-regions-) were observed in more than one patients (recurrent; see Supplementary Material, Table S10). We then filtered these CNVs against the DGV and this resulted in 114 CNVs that not only were exclusive to ARM patients, but also were absent in control individuals in the DGV. While nine were recurrent (distributed in four CNVR-regions-), 105 were non-recurrent (see Supplementary Material, Table S11). These CNVs were subsequently classified according to their genic content (genic and non-genic CNVs).

Gene-based analysis: genes of the WNT and SHH signaling pathways are disrupted in ARM patients

As the pathogenicity of a genic CNV may be linked to not only the number of genes included but also to the biological plausibility of the gene in relation to the phenotype under study, genes intersected by the CNVs were carefully scrutinized and prioritized. We then performed a gene-based analysis in which all CNVs were included. We identified 496 genes that were not disrupted in controls (472 unique—disrupted in one patient; 24 recurrent, see Supplementary Material, Table S12). After filtering against the CNVs of 11 943 healthy controls from the DGV, 79 genes (see Supplementary Material, Table S13) were found to be uniquely interfered by CNVs in our patients (see Supplementary Material, Fig. S7).

Given the relevance of the *WNT* and *SHH* interrelated signaling pathways in embryonic development, together with

the human ARM-reminiscent phenotypes displayed by animal models for those pathways (21,22), we explored those CNV events that overlap with gene members of the *WNT/SHH* signaling pathways. We identified two patients (MG-IA349C—isolated; MG-IA78C with Down syndrome) with a 34.4 kb heterozygous deletion spanning exon 5 to exon 8 of *INTU* (4q28.1; Fig. 1A), and one patient (MG-IA147C: isolated imperforate anus and diagnosed with autism at the age of 6) with a duplication (three copies) of the whole *DKK4* gene (8p11.21; Fig. 1B). These computationally predicted CNVs were validated using Taqman® copy number assays (Fig. 2A and B). Importantly, the duplication of *DKK4* was *de novo* (Fig. 2C). The inherited or *de novo* nature of the *INTU* deletion could not be established as parental DNA was not available.

INTU (inturned planar cell polarity effector homolog) encodes a structural protein that controls ciliogenesis and the organization of the cytoskeleton (governing the apical actin assembly and controlling the orientation of ciliary microtubules) and its disruption is associated with the failure in planar cell polarity (PCP) and hedgehog signaling pathways (23,24). Many hedgehog pathway components, including the *Gli* family of transcription factors, localize to cilia and proper *Intu* expression is required for their ciliary translocation to the nucleus (25–28). Mutations in *Intu* cause loss of *Shh* signaling (*Gli1* protein) in the mouse posterior spinal cord, and mice die at E9.5 (25). Importantly, defects in *SHH* (i.e. mutation in *Gli2* and/or *Gli3*) or PCP signaling lead to the ARM phenotype in mice (21,22). The effect is dosage-dependent, i.e. more severe phenotypes are observed when two copies of the mutated genes are defective (21). Yet no coding region mutations in these genes have been identified in humans affected with isolated ARMs. *GLI3* coding sequence mutations are associated with Pallister–Hall syndrome (OMIM #146510) which includes imperforate anus its phenotypic spectrum.

Excess of DKK4 leads to ARMs

DKK4 encodes a secreted protein member of dickkopf (*DKK*) family of *WNT* regulators. *DKKs*, together with *WNT* secreted proteins play an important role in antero-posterior axial patterning, limb development, somitogenesis and eye formation (29). During development, *DKK4* competes with *WNT* ligands for the co-receptors, thus antagonizing *WNT* signaling

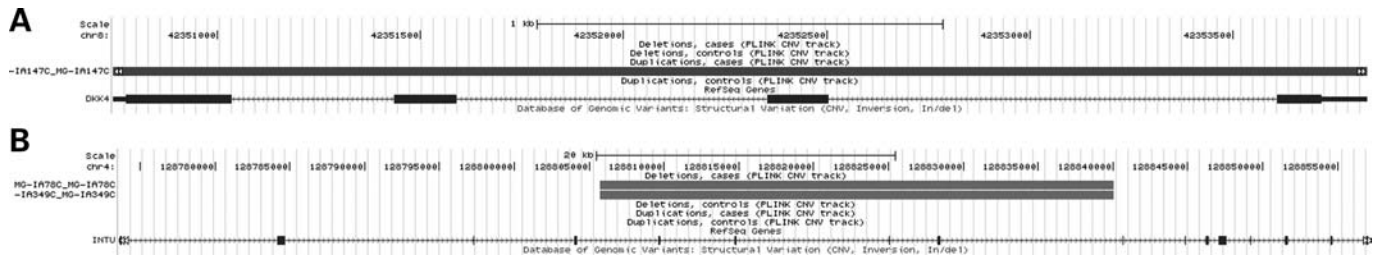


Figure 1. UCSC Genome Browser showing the two ARMs implicated genes: (A) *DKK4* and (B) *INTU*, and the CNVs interfering with the gene regions. (A) UCSC Genome Browser showing the gene region of *DKK4* and the part of the 1.32 Mb duplication (chromosome 8p12–q11.21, three copies) observed in an isolated ARM patient (MG-IA147C). The duplication is represented by the blue bar. There is no CNV observed in the normal samples submitted to the DGV. (B) UCSC Genome Browser showing the gene region of *INTU* and the 34 kb hemizygous deletions in 4q28.1 observed in two ARM patients (MG-IA349C with isolated-ARM; MG-IA78C with Down syndrome). The deletions are represented by the red bars. There is no CNV observed in the normal samples submitted to the DGV.

pathway. In mice, defects in *Wnt* signaling pathway lead to ARMs (12,13,29,30).

From all of the above, it would appear that deregulation of the *WNT* pathway by overexpression of *DKK4* may further impair *WNT* signaling and lead to ARMs. We then tested this hypothesis in a mouse anorectum organotypic culture. The urogenital sinus and the hindgut are connected at the cloaca at E12 in mouse embryo (Fig. 3A). By E13.5, the cloaca is being separated by a sheet of mesenchyme called urorectal septum, which has elongated and descended towards the cloaca membrane (Fig. 3B), and at the same time, the genital tubercle has grown distally due to the proliferation of the rostral mesoderm of the genital tubercle. This process compartmentalizes the cloaca into two cavities from which the anal opening and urethra opening will originate, respectively. In control culture, the genital tubercle has grown distally after 36 h. The urorectal septum has already elongated and reached the cloaca membrane (Fig. 3C). In contrast, treatment with *Dkk4* protein (Fig. 3D) perturbed the growth of the urorectal septum and resulted in the lack of cloaca compartmentalization. The hollow space resembled the phenotype of persistent cloaca as shown in the mid-sagittal section depicted in Figure 3B. However, the distal growth of the genital tubercle appeared unaffected by the addition of *Dkk4* protein. This experiment proves that excess of *DKK4* may lead to ARMs. Therefore, it would appear that *DKK4* is a candidate gene for ARMs.

However, as *Intu* is a cytoplasmatic protein, we could not test the effect of deletion directly by employing the same experiment design. Mutations in *Intu* had been reported before that they cause loss of *SHH* signaling (*Gli1* protein) in the mouse posterior spinal cord, and mice die at E9.5 (25). Therefore, remaining support for selecting *INTU* as a possible candidate gene is the fact that it is within a rare CNV and that it is involved in *SHH* signaling.

DISCUSSION

CNVs are abundant and can be functionally influential. Their importance in human diseases has become increasingly apparent over the past 5 years. With the advancement in detection resolution and genome coverage of genotyping arrays, detection of CNVs at genome-wide scale is possible. Based on the intensity of SNP and CNV probes in the array, CNVs

can be predicted and analyzed for their association to the disease. Several large-scale studies have reported that CNVs, especially rare CNVs, may account for a significant proportion of human phenotypic variation, including disease susceptibility (31,32). Data from the latest CNV studies indicate that disease status is more likely to be caused by an accumulation of rare CNVs rather than by differences in CNP loads (32). We now appreciate that at least 15% of human neurodevelopmental diseases are due to rare and large copy number changes which lead to local dosage imbalance for dozens of genes. Large CNVs, both inherited and *de novo*, have been implicated in the etiology of autism, schizophrenia, kidney dysfunction and congenital heart disease (33). Studies of the general population suggest that collectively, rare CNVs are quite common and are under strong purifying selection. This implies that a significant fraction of the human population carries an unbalanced genome and such individuals, may be sensitized by the effect of another variant interacting with these CNVs in a digenic manner.

One of the major challenges in CNV discovery is to discriminate between benign and pathological variants. The rarer or longer the CNV, the more likely it is to be pathogenic. CNV is also more likely to be pathogenic when the genetic event is *de novo*, when CNVs are found only in patients and when the genes encompassed or disrupted by the CNV belong to a pathway known to contain genes associated with a similar phenotype under study (34,35). Indeed, many of the CNVs identified in this study meet some of the above criteria. Besides the 114 rare CNVs exclusive to ARM patients, rare CNVs were overall in excess in the patients. Moreover, some CNVs not only intersected with gene members of pathways (i.e. *SHH* and *WNT*) that are involved in the development of the anorectal region, but also contained genes associated with similar or related phenotypes in mice and humans (12,13,21,30,36,37). Importantly, we could prove that the *DKK4* duplication was *de novo*. Interestingly, while mice mutant for *SHH* gene members (*Shh*, *Gli2*, or *Gli3*) displayed congenital defects that include ARMs as a common feature, point mutations in the human orthologs (*SHH*, *GLI3*) are associated with syndromes or genetically heterogeneous disorders in which the ARM phenotype is not always the norm.

As any other developmental disorders, rare chromosome aberrations (CNV longer than 1 Mb) have been reported in 4.5–11% of the patients, mostly with syndromic ARMs

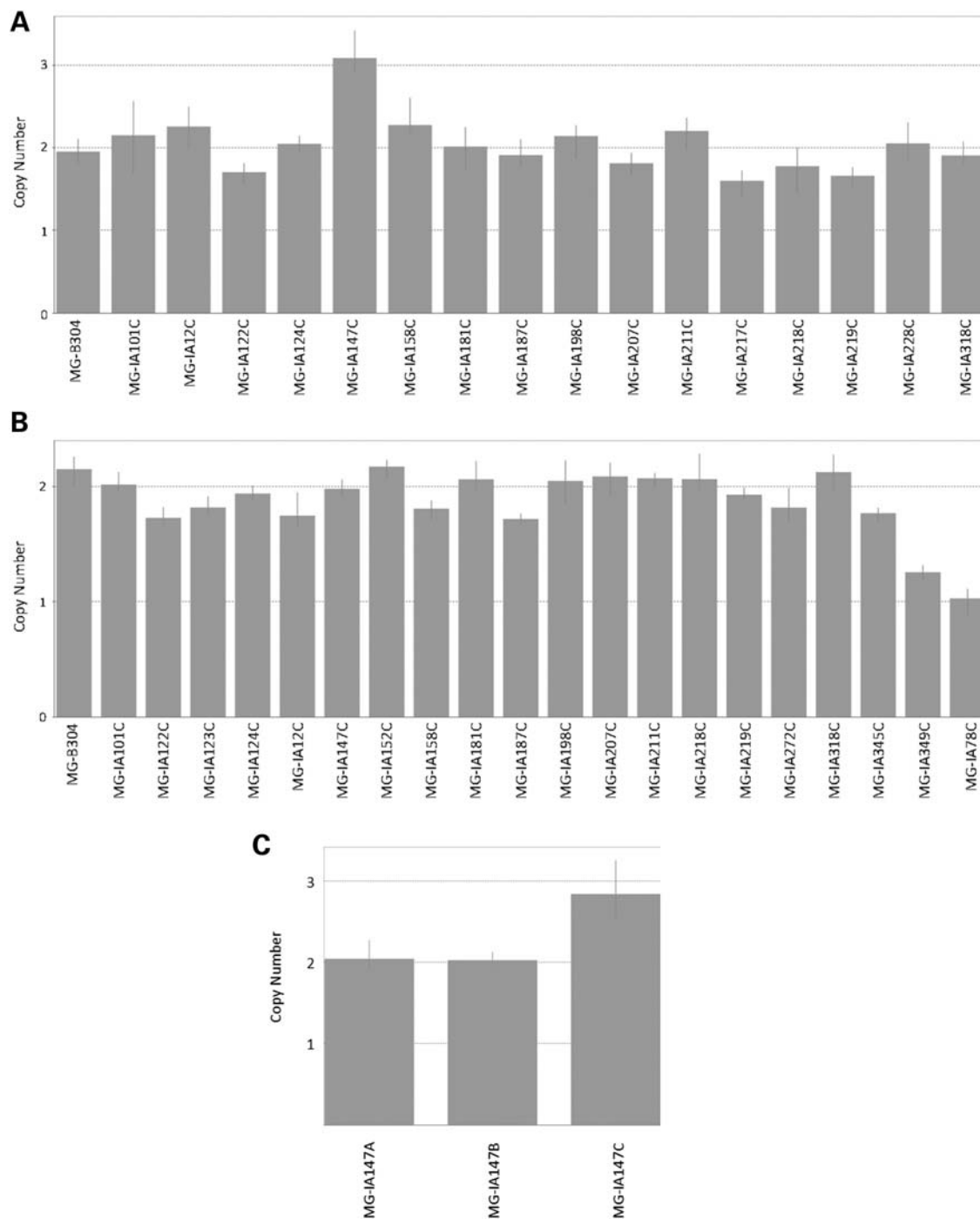


Figure 2. Validation of CNVs interfering with ARMs implicated genes: (A) *DKK4* and (B) *INTU*. (A) Validation of the duplication (three copies) spanning the whole *DKK4* gene in one ARMs case (MG-IA147C) and of the normal copy number (two copies) in other GWAS ARMs cases. MG-IA147C (the sixth bar from the left) has three *DKK4* copies while the rest of samples tested had two copies. (B) Validation of the deletion (one copy) spanning *INTU* in two ARMs cases (MG-IA78C and MG-IA349C) and of the normal copy number (two copies) in other GWAS ARMs cases. MG-IA78C and MG-IA349C (the two rightmost bars) have one copy while other GWAS subjects had two copies. (C) Validation of the duplication (three copies) spanning the whole *DKK4* gene in one ARMs case (MG-IA147C) and proving it to be a *de novo* event by validating the normal copy number in the parents (MG-IA147A and MG-IA147B) of this ARMs case.

(20). Indeed, ARMs can be part of the phenotypic spectrum of many chromosomal anomalies such as trisomy 13, 18, 21 or 22 to mention a few (20,38). Here, we identified 12 chromosomal aberrations (besides trisomy 21) that were unique to isolated ARM patients, indicating a role for those aberrations in isolated ARMs.

Developmental disorders are notoriously associated with a myriad of rare chromosomal aberrations and CNVs, and their rarity makes clinical interpretation problematic and genotype–phenotype correlations uncertain. A genomic rearrangement shared by patients with phenotypic features in common surely implies greater certainty in the pathogenic nature of

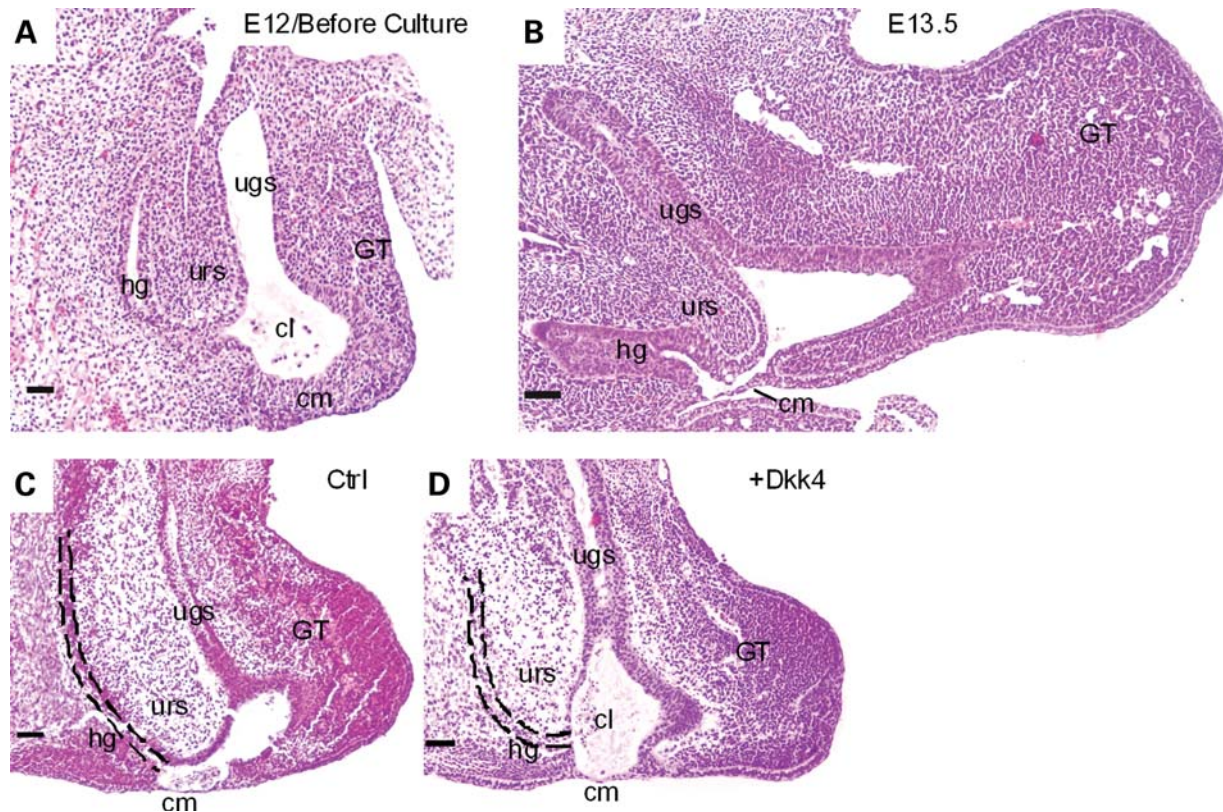


Figure 3. Excess of Dkk4 protein led to ARMs as shown by mouse embryonic anorectum culture. Mid-sagittal sections of the anorectums of E12 (A) and E13.5 (B) ICR mouse embryos were shown. Organ culture of E12 anorectums from ICR mice treated for 36 h in culture with (C) 1% BSA as control (Ctrl), and (D) 1.5 mg/ml of Dkk4 protein (+Dkk4) were processed and sectioned. Mid-sagittal sections of cultured anorectums were shown. In control culture, the genital tubercle has grown distally after 36 h. The urorectal septum has elongated and reached the cloaca membrane (C). In contrast, treatment with Dkk4 protein (D) perturbed the growth of the urorectal septum and resulted in the lack of cloaca compartmentalization. The hollow space resembled the phenotype of persistent cloaca as shown in the mid-sagittal section depicted in Figure 3B. However, the distal growth of the genital tubercle appeared unaffected by the addition of Dkk4 protein. cl, cloaca; cm, cloaca membrane; hg, hindgut; GT, genital tubercle; ugs, urogenital sinus; urs, urorectal septum. Bar 0.05 mm. Dotted lines demarcate the hindgut.

the CNV. Comparison of the chromosomal aberrations unique to ARM patients with those reported in patients of DECIPHER (Table 2) or in The European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA) has revealed the presence of the 22q11.2 deletion in the later of six cases with syndromic ARMs (20). A similar deletion was detected in one of our patients with isolated ARMs. Likewise, our patients shared rearrangements with DECIPHER patients with related phenotypes.

Surely, rare CNVs implicate several novel disease candidates genes since there is a multitude of ways in which gene function can be altered by these structural variations (alter gene dosage, disrupt coding sequences or affect gene regulation and consequently may lead to disease). Indeed, our initial analyses showed an excess of rare CNVs in ARM patients when compared with controls, with 79 genes disrupted by CNVs uniquely in patients, providing a wealth of putative disease candidate genes, in particular *DKK4* and *INTU*. Common genetic variants (SNPs and CNPs) have been found to have little contribution to the condition as ARMs, being phenotypically heterogeneous, are likely to result from rare mutations in a variety of genes. As in many other congenital diseases, several genes acting in different tissues and at different developmental stages may be involved in ARMs.

Mutations in any of these genes could lead to the phenotype. Because each gene and its product are subject to complex regulation at every stage, the reach of a mutational event will depend on the gene implicated. Thus, the complexity of these molecular events would explain both the genetic heterogeneity and phenotypic variability of the condition. Our data suggest that the condition is likely caused by rare variants (CNV or single point mutation) in any of the genes implicated in the developmental processes. This would be in line with the lack of association signals for common genetic variants and the manifestation of the disease. Thus, rare DNA variations in any of the developmental genes implicated could not only lead to the phenotype but also explain its variability on the grounds mentioned earlier.

MATERIALS AND METHODS

Subjects and ethics statement

The overall study was approved by the institutional review board of The University of Hong Kong together with the Hospital Authority (IRB: UW 07-321). Blood samples were drawn from all participants after obtaining informed consent (parental consent in newborns and children below age 7).

ARM patients

A total of 363 Chinese sporadic ARM patients (isolated or with additional associated anomalies) had prospectively been collected throughout Hong Kong and Mainland China. All patients included in this study went through renal ultrasound, lumbosacral radiography and ECHO cardiography. Patients were initially grouped into discovery phase by genome-wide scan (185 patients) and replication series (178 non-syndromic patients). The overall male-to-female ratio was $\sim 1.4:1$. Phenotypic characteristics of the patients are summarized in Supplementary Material, Table S1. Patients were defined as syndromic if associated anomalies were observed in addition to ARMs (see Supplementary Material, Table S1b). In the discovery phase, we included 46 syndromic ARM patients, among whom 15 had Down syndrome.

Controls

As controls, we used the DNA sample from a total of 3249 Chinese individuals (discovery phase: $n = 3072$, replication phase: $n = 177$) whom were also recruited throughout Hong Kong and Mainland China. For the discovery phase, we included 3072 individuals who were either phenotypically normal ($n = 1421$) or affected with conditions other than ARMs ($n = 1651$). These 1651 individuals ('shared' controls) had been included in other GWAS conducted in our institution [i.e. patients affected with schizophrenia, hypertension (39), epilepsy (40) or systemic lupus erythematosus (41)]. Details on the characteristics of the shared controls can be found in Supplementary Material, Table S2. Individuals affected with other conditions were used as controls because: (i) disease-specific effect in the controls can be diluted if it consists of balanced disease samples; (ii) sharing samples from different projects can detect differential errors due to different DNA preparation and genotyping; (iii) cost is reduced for collecting phenotype and genotype data from additional control samples; (iv) power increases with the number of controls used. For the SNP replication phase, 177 phenotypically normal individuals were recruited as controls. For the CNV analysis, we included 868 individuals who are phenotypically normal from other studies (111 controls from hypertension study and 757 individuals from osteoporosis study) (42).

Discovery phase

Whole-genome scan

The whole-genome scan was performed at deCODE Genetics (Reykjavik, Iceland) using Illumina Human 610-Quad Bead-Chips which assay 599 011 SNPs across the genome and 21 890 intensity-only CNV probes. SNP calls were provided by deCODE. SNP quality control and association tests together with the results are detailed in the supplementary Material and Methods.

CNVs: predictions

CNV segments were predicted by two programs, PennCNV (43) and QuantiSNP (44), the two most efficient and publicly available CNV calling algorithms for Illumina data (45). Both programs implement hidden Markov models (HMM) while PennCNV integrate additional information in CNV prediction

(i.e. population allele frequency and distance between adjacent SNPs) when compared with QuantiSNP.

CNVs: quality controls

In spite of the advancement in CNVs detection using genome-wide SNP arrays and better CNV prediction algorithms, the concordance of CNVs called by different algorithms is still low ($< 50\%$) (46). This implies a high false positive rate in CNV predictions. To obtain high-confidence calls, we only used the overlapping region of CNVs called by PennCNV and QuantiSNP. Before selecting the overlapping CNV regions, quality controls were done separately for the CNV predicted by two programs.

For both PennCNV and QuantiSNP callings, CNVs shorter than 1 kb or called with fewer than three probes were removed. In addition to these filtering criteria, we also remove CNVs with maximum Bayes factor < 10 for the predictions by QuantiSNP. In the analysis, only those regions intersected by CNVs called by both programs were included. Samples with genome-wide LRR standard deviation > 3.5 or with more than 500 CNVs called were excluded from the analysis (patients = 5; controls = 17).

CNVs might be artificially split by either of the calling programs. To circumvent this issue, adjacent CNVs of the same type (i.e. duplication or deletion) were merged if the length of gap in between was shorter than half of total length of the two consecutive CNV segments.

After quality control, 170 ARM cases (northern Chinese: 98, southern Chinese: 72) and 851 controls (northern Chinese: 37, southern Chinese: 784) with 4129 and 21 027 CNVs, respectively, in total were analyzed for the discovery of disease-associated CNV regions.

CNVs analysis

Common CNPs analysis, CNV replication and CNV validation are detailed in the Supplementary material online.

Rare CNVs. Rare CNVs are defined as CNVs that are observed in $< 1\%$ of samples in the data set (i.e. observed in less than or equal to 10 samples in this study). We first compared the global burden of rare CNVs between cases and controls. Then, we defined rare CNV regions (rare CNVRs; see Supplementary material online) and analyzed each of them individually.

Global burden. Global burden tests were performed in terms of CNV length, number of CNVs and genes overlapped. Permutations tests conducted by PLINK were used to determine the statistical significance (1 000 000 permutations for each burden test). The global burden tests were used to examine the possible differences in terms of common or rare CNVs enrichment between (i) ARM patients and controls (empirical one-sided P -values are reported); and (ii) northern Chinese controls and southern Chinese controls (empirical two-sided P -values are reported). The enrichment of long CNVs (defined as those CNVs with length longer than 1 Mb) between ARM patients and controls was also examined (empirical one-sided P -values are reported).

Gene-based CNV analysis

With the overlapped CNV calls from pennCNV and quantiSNP, we selected those gene regions that were only found to be disrupted in the ARMs cases, but not in the 868 controls or the 11 943 unique normal individuals from the DGV. We then examined the developmental genes that were only disrupted in ARMs cases.

Mouse anorectum organotypic culture

Timed pregnant mice (strain ICR) at embryonic day E12 were sacrificed. The embryonic anorectums were treated and cultured as described previously (47). Different treatments were applied to simulate different conditions: (i) as control: control culture was treated with PBS containing 0.1%BSA; (ii) excess of *Dkk4* protein (secreted protein): recombinant mouse *Dkk4* (R&D) proteins were added to the culture medium at a concentration of 1.5 µg/ml.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the research grant from the Hong Kong Research Grants Council (HKU 775608M to P.K.-H.T.) and the HKU seed funding programme for basic research (200911159060 to V.C.-H.L.). Support was also received from the University of Hong Kong Strategic Research Theme on Genomics.

REFERENCES

- Cuschieri, A. (2001) Descriptive epidemiology of isolated anal anomalies: a survey of 4.6 million births in Europe. *Am. J. Med. Genet.*, **103**, 207–215.
- Kosho, T., Nakamura, T., Kawame, H., Baba, A., Tamura, M. and Fukushima, Y. (2006) Neonatal management of trisomy 18: clinical details of 24 patients receiving intensive treatment. *Am. J. Med. Genet. A.*, **140**, 937–944.
- Lin, H.Y., Lin, S.P., Chen, Y.J., Hung, H.Y., Kao, H.A., Hsu, C.H., Chen, M.R., Chang, J.H., Ho, C.S., Huang, F.Y. *et al.* (2006) Clinical characteristics and survival of trisomy 18 in a medical center in Taipei, 1988–2004. *Am. J. Med. Genet. A.*, **140**, 945–951.
- de Buys Roessingh, A.S., Mueller, C., Wiesenauer, C., Bensoussan, A.L. and Beauoyer, M. (2009) Anorectal malformation and Down's syndrome in monozygotic twins. *J. Pediatr. Surg.*, **44**, e13–e16.
- Levitt, M.A. and Pena, A. (2007) Anorectal malformations. *Orphanet J. Rare Dis.*, **2**, 33.
- Stoll, C., Alembik, Y., Dott, B. and Roth, M.P. (2007) Associated malformations in patients with anorectal anomalies. *Eur. J. Med. Genet.*, **50**, 281–290.
- Schramm, C., Draaken, M., Tewes, G., Bartels, E., Schmiedede, E., Marzheuser, S., Grasshoff-Derr, S., Hosie, S., Holland-Cunz, S., Priebe, L. *et al.* (2011) Autosomal-dominant non-syndromic anal atresia: sequencing of candidate genes, array-based molecular karyotyping, and review of the literature. *Eur. J. Pediatr.*, **170**, 741–746.
- Weinstein, E.D. (1965) Sex-linked imperforate anus. *Pediatrics*, **35**, 715–718.
- Vangelder, D.W. and Kloepfer, H.W. (1961) Familial anorectal anomalies. *Pediatrics*, **27**, 334–336.
- Schwoebel, M.G., Hirsig, J., Schinzel, A. and Stauffer, U.G. (1984) Familial incidence of congenital anorectal anomalies. *J. Pediatr. Surg.*, **19**, 179–182.
- Stoll, C., Alembik, Y., Roth, M.P. and Dott, B. (1997) Risk factors in congenital anal atresias. *Ann. Genet.*, **40**, 197–204.
- Jia, H., Chen, Q., Zhang, T., Bai, Y., Yuan, Z. and Wang, W. (2011) Wnt5a expression in the hindgut of fetal rats with chemically induced anorectal malformations—studies in the ETU rat model. *Int. J. Colorectal Dis.*, **26**, 493–499.
- Tai, C.C., Sala, F.G., Ford, H.R., Wang, K.S., Li, C., Minoo, P., Grikscheit, T.C. and Bellusci, S. (2009) Wnt5a knock-out mouse as a new model of anorectal malformation. *J. Surg. Res.*, **156**, 278–282.
- McCarroll, S.A. and Altshuler, D.M. (2007) Copy-number variation and association studies of human disease. *Nat. Genet.*, **39**, S37–S42.
- Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., Giannoulatou, E. *et al.* (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, **464**, 713–720.
- Schramm, C., Draaken, M., Bartels, E., Boemers, T.M., Aretz, S., Brockschmidt, F.F., Nothen, M.M., Ludwig, M. and Reutter, H. (2011) *De novo* microduplication at 22q11.21 in a patient with VACTERL association. *Eur. J. Med. Genet.*, **54**, 9–13.
- Yamagishi, H., Ishii, C., Maeda, J., Kojima, Y., Matsuoka, R., Kimura, M., Takao, A., Momma, K. and Matsuo, N. (1998) Phenotypic discordance in monozygotic twins with 22q11.2 deletion. *Am. J. Med. Genet.*, **78**, 319–321.
- Worthington, S., Colley, A., Fagan, K., Dai, K. and Lipson, A.H. (1997) Anal anomalies: an uncommon feature of velocardiofacial (Shprintzen) syndrome? *J. Med. Genet.*, **34**, 79–82.
- Schulze, B.R., Tariverdian, G., Komposch, G. and Stellzig, A. (2001) Misclassification risk of patients with bilateral cleft lip and palate and manifestations of median facial dysplasia: a new variant of del(22q11.2) syndrome? *Am. J. Med. Genet.*, **99**, 280–285.
- Marcelis, C., de Blaauw, I. and Brunner, H. (2011) Chromosomal anomalies in the etiology of anorectal malformations: a review. *Am. J. Med. Genet. A.*, **155**, 2692–2704.
- Mo, R., Kim, J.H., Zhang, J., Chiang, C., Hui, C.C. and Kim, P.C. (2001) Anorectal malformations caused by defects in sonic hedgehog signaling. *Am. J. Pathol.*, **159**, 765–774.
- Wen, J., Chiang, Y.J., Gao, C., Xue, H., Xu, J., Ning, Y., Hodes, R.J., Gao, X. and Chen, Y.G. (2010) Loss of Dact1 disrupts planar cell polarity signaling by altering dishevelled activity and leads to posterior malformation in mice. *J. Biol. Chem.*, **285**, 11023–11030.
- Park, T.J., Haigo, S.L. and Wallingford, J.B. (2006) Ciliogenesis defects in embryos lacking inturned or fuzzy function are associated with failure of planar cell polarity and Hedgehog signaling. *Nat. Genet.*, **38**, 303–311.
- Murdoch, J.N. and Copp, A.J. (2010) The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. *Birth Defects Res. A. Clin. Mol. Teratol.*, **88**, 633–652.
- Zeng, H., Hoover, A.N. and Liu, A. (2010) PCP effector gene Inturned is an important regulator of cilia formation and embryonic development in mammals. *Dev. Biol.*, **339**, 418–428.
- Kim, J., Kato, M. and Beachy, P.A. (2009) Gli2 trafficking links Hedgehog-dependent activation of Smoothed in the primary cilium to transcriptional activation in the nucleus. *Proc. Natl Acad. Sci. USA*, **106**, 21666–21671.
- Goetz, S.C. and Anderson, K.V. (2010) The primary cilium: a signalling centre during vertebrate development. *Nat. Rev. Genet.*, **11**, 331–344.
- Goetz, S.C., Ocbina, P.J. and Anderson, K.V. (2009) The primary cilium as a Hedgehog signal transduction machine. *Methods Cell Biol.*, **94**, 199–222.
- Niehrs, C. (2006) Function and biological roles of the Dickkopf family of Wnt modulators. *Oncogene*, **25**, 7469–7481.
- Nakata, M., Takada, Y., Hishiki, T., Saito, T., Terui, K., Sato, Y., Koseki, H. and Yoshida, H. (2009) Induction of Wnt5a-expressing mesenchymal cells adjacent to the cloacal plate is an essential process for its proximodistal elongation and subsequent anorectal development. *Pediatr. Res.*, **66**, 149–154.
- Feuk, L., Carson, A.R. and Scherer, S.W. (2006) Structural variation in the human genome. *Nat. Rev. Genet.*, **7**, 85–97.
- Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P. *et al.* (2010) Origins and functional impact of copy number variation in the human genome. *Nature*, **464**, 704–712.

33. Girirajan, S., Campbell, C.D. and Eichler, E.E. (2011) Human copy number variation and complex genetic disease. *Ann. Rev. Genet.*, **45**, 203–226.
34. Lee, C., Iafrate, A.J. and Brothman, A.R. (2007) Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat. Genet.*, **39**, S48–S54.
35. Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P., Hurler, M.E. and Feuk, L. (2007) Challenges and standards in integrating surveys of structural variation. *Nat. Genet.*, **39**, S7–15.
36. Garcia-Barcelo, M.M., Chi-Hang Lui, V., Miao, X., So, M.T., Yuk-yu Leon, T., Yuan, Z.W., Li, L., Liu, L., Wang, B., Sun, X.B. *et al.* (2008) Mutational analysis of SHH and GLI3 in anorectal malformations. *Birth Defects Res. A. Clin. Mol. Teratol.*, **82**, 644–648.
37. Ulloa, F. and Marti, E. (2010) Wnt won the war: antagonistic role of Wnt over Shh controls dorso-ventral patterning of the vertebrate neural tube. *Dev Dyn*, **239**, 69–76.
38. Cuschieri, A. (2002) Anorectal anomalies associated with or as part of other anomalies. *Am. J. Med. Genet.*, **110**, 122–130.
39. Guo, Y., Tomlinson, B., Chu, T., Fang, Y.J., Gui, H., Tang, C.S., Yip, B.H., Cherny, S.S., Hur, Y.M., Sham, P.C. *et al.* (2012) A genome-wide linkage and association scan reveals novel loci for hypertension and blood pressure traits. *PLoS One*, **7**, e31489.
40. Guo, Y., Baum, L.W., Sham, P.C., Wong, V., Ng, P.W., Lui, C.H., Sin, N.C., Tsoi, T.H., Tang, C.S., Kwan, J.S. *et al.* (2012) Two-stage genome-wide association study identifies variants in CAMSAP1L1 as susceptibility loci for epilepsy in Chinese. *Hum. Mol. Genet.*, **21**, 1184–1189.
41. Yang, W., Shen, N., Ye, D.Q., Liu, Q., Zhang, Y., Qian, X.X., Hirankarn, N., Ying, D., Pan, H.F., Mok, C.C. *et al.* (2010) Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet.*, **6**, e1000841.
42. Kung, A.W., Xiao, S.M., Cherny, S., Li, G.H., Gao, Y., Tso, G., Lau, K.S., Luk, K.D., Liu, J.M., Cui, B. *et al.* (2010) Association of JAG1 with bone mineral density and osteoporotic fractures: a genome-wide association study and follow-up replication studies. *Am. J. Hum. Genet.*, **86**, 229–239.
43. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H. and Bucan, M. (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.*, **17**, 1665–1674.
44. Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C. and Ragoussis, J. (2007) QuantiSNP: an objective Bayes hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.*, **35**, 2013–2025.
45. Dellinger, A.E., Saw, S.M., Goh, L.K., Seielstad, M., Young, T.L. and Li, Y.J. (2010) Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res.*, **38**, e105.
46. Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A.C., Thiruvahindrapuram, B., Macdonald, J.R., Mills, R. *et al.* (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat. Biotechnol.*, **29**, 512–520.
47. Ma, L.M., Wang, Z., Wang, H., Li, R.S., Zhou, J., Liu, B.C. and Baskin, L.S. (2009) Estrogen effects on fetal penile and urethral development in organotypic mouse genital tubercle culture. *J. Urol.*, **182**, 2511–2517.