AGORA, a Data- and Biobank for Birth Defects and Childhood Cancer

Iris A.L.M. van Rooij*[1], Loes F.M. van der Zanden[1], Ernie M.H.F. Bongers[2], Kirsten Y. Renkema[3], Charlotte H.W. Wijers[1], Michelle Thonissen[4], Elisabeth M.J. Dokter[1,5], Carlo L.M. Marcelis[2], Ivo de Blaauw[6,7], Marc H.W.A. Wijnen[6], Peter M. Hoogerbrugge[8], Jos P.M. Bokkerink[8], Michiel F. Schreuder[8], Linda Koster-Kamphuis[8], Elisabeth A.M. Cornelissen[8], Livia Kapusta[8], Arno F.J. van Heijst[8], Kian D. Liem[8], Robert P.E. de Gier[9], Anne Marie Kuijpers-Jagtman[4], Ronald J.C. Admiraal[10], Stefaan J. Bergé[11], Jan Jaap van der Biezen[5], An Verdonck[12], Vincent Vander Poorten[13], Greet Hens[13], Jasmien Roosenboom[14], Marc R. Lilien[15], Tom P. de Jong[16,17], Paul Broens[18], Rene Wijnen[7], Alice Brooks[19], Barbara Franke[2,20], Han G. Brunner[2], Carine E.L. Carels[4,12], Nine V.A.M. Knoers[3], Wout F.J. Feitz[9], and Nel Roeleveld[1,8]

BACKGROUND: Research regarding the etiology of birth defects and childhood cancer is essential to develop preventive measures, but often requires large study populations. Therefore, we established the AGORA data- and biobank in the Netherlands. In this study, we describe its rationale, design, and ongoing data collection. METHODS: Children diagnosed with and/or treated for a structural birth defect or childhood cancer and their parents are invited to participate in the AGORA data- and biobank. Controls are recruited through random sampling from municipal registries. The parents receive questionnaires about demographics, family and pregnancy history, health status, prescribed medication, lifestyle, and occupational exposures before and during the index pregnancy. In addition, blood or saliva is collected from children and parents, while medical records are reviewed for diagnostic information. RESULTS: So far, we have collected data from over 6,860 families (3,747 birth defects, 905 childhood cancers, and 2,208 controls). The types of birth defects vary widely and comprise malformations of the digestive, respiratory, and urogenital tracts as well as facial, cardiovascular, kidney, skeletal, and central nervous system anomalies. The most frequently occurring childhood cancer types are acute lymphatic leukemia, Hodgkin and non-Hodgkin lymphoma, Wilms' tumor, and brain and spinal cord tumors. Our genetic and/or epidemiologic studies have been focused on hypospadias, anorectal malformations, congenital anomalies of the kidney and urinary tract (CAKUT), and orofacial clefts. CONCLUSION: The large AGORA data- and biobank offers great opportunities for investigating genetic and nongenetic risk factors for disorders in children and is open to collaborative initiatives.

Birth Defects Research (Part A) 106:675–684, 2016.

Key words: etiology; cancer; congenital malformations; environment; genetics; risk factor

Supported by the Radboud university medical center.
[1]Department for Health Evidence, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, The Netherlands
[2]Department of Human Genetics, Radboud university medical center, Nijmegen, The Netherlands
[3]Department of Genetics, University Medical Center Utrecht, The Netherlands
[4]Department of Orthodontics and Craniofacial Biology, Radboud university medical center, Nijmegen, The Netherlands
[5]Department of Plastic Surgery and Hand Surgery, Medical Centre Leeuwarden, Leeuwarden, the Netherlands
[6]Department of Surgery, Pediatric Surgery, Radboudumc Amalia Children's Hospital, Nijmegen, The Netherlands
[7]Department of Pediatric Surgery, Erasmus Medical Center, Rotterdam, The Netherlands
[8]Department of Pediatrics, Radboudumc Amalia Children's Hospital, Nijmegen, The Netherlands
[9]Department of Urology, Pediatric Urology, Radboudumc Amalia Children's Hospital, Nijmegen, The Netherlands
[10]Department of Otorhinolaryngology, Radboud university medical center, Nijmegen, The Netherlands
[11]Department of Oral and Maxillofacial surgery, Radboud university medical center, Nijmegen, The Netherlands
[12]Department of Oral Health Sciences-Orthodontics and Leuven Cleft Lip and Palate Team, KU Leuven, University Hospitals KU Leuven, Belgium
[13]Department of Otorhinolaryngology, Head and Neck Surgery and Leuven Cleft Lip and Palate Team, University Hospitals KU Leuven, Belgium
[14]Department of Neurosciences, Experimental Otorhinolaryngology, KU Leuven, Belgium
[15]Department of Pediatric Nephrology, Wilhelmina Children's Hospital, University Medical Center Utrecht, The Netherlands
[16]Department of Pediatric Urology, Wilhelmina Children's Hospital, University Medical Center Utrecht, The Netherland
[17]Department of Pediatric Urology, Emma Children's Hospital, Academic Medical Center Amsterdam, The Netherlands
[18]Department of Surgery, Division of Pediatric Surgery, University Medical Center Groningen, Groningen, The Netherlands
[19]Department of Clinical Genetics, Erasmus Medical Centre, Rotterdam, The Netherlands
[20]Department of Psychiatry, Radboud university medical center, Nijmegen, The Netherlands
Drs. van Rooij and van der Zanden contributed equally to this work.
*Correspondence to: Iris A.L.M. van Rooij, Department for Health Evidence (133), Radboud university medical center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands. E-mail: iris.vanrooij@radboudumc.nl

Published online 6 May 2016 in Wiley Online Library (wileyonlinelibrary.com). Doi: 10.1002/bdra.23512

Introduction

Approximately 1 in 40 children is born with a major structural birth defect, which has an enormous impact on the lives of these children and their parents, because of associated morbidity, long-term disability, and mortality (Corsello and Giuffre, 2012). Due to numerous medical procedures, treatments, and visits to specialists, the burden on society is also substantial. The etiology is known for only a small percentage of structural birth defects. This mostly applies to syndromes caused by monogenic or chromosomal abnormalities, such as Noonan and Down syndrome, and to syndromes and defects caused by major teratogenic exposures, such as fetal alcohol syndrome and the consequences of thalidomide exposure. The etiology of the majority of birth defects, however, is unknown and suggested to be multifactorial, involving a complex interplay of genetic and nongenetic factors, such as environmental exposures, medical drug use, maternal illness, or nutritional imbalances. Both genetic and nongenetic factors may interfere with regulatory molecular pathways in critical periods of embryogenesis and organogenesis.

The incidence rate of cancer among children under 15 years of age is approximately 1 in 7,150 (Kaatsch, 2010). Treatment is often extensive and has an enormous impact on these children and their surroundings in physical, mental, and social respects. Although the probability of survival increased considerably in the past decades, childhood cancer is still the second most common cause of childhood mortality in developed countries (Kaatsch, 2010). Like most cancers, childhood cancers result from multiple molecular changes. The etiology of the majority of childhood cancers is believed to be multifactorial (Buka et al., 2007), and the initial predisposing event is thought to occur before birth in most cases. This *in utero* initiation hypothesis is circumstantially supported by the young age at diagnosis, by the resemblance of some childhood cancer cells to cells found during fetal development, and by case reports describing prenatal diagnosis of several cancer types (Stiller, 2004; Spector et al., 2008; Spector, 2010).

Major obstacles in investigating the etiology of birth defects and childhood cancer are the large patient groups needed in combination with the rareness of most of these disorders. A genome-wide association study (GWAS), for instance, requires a sample size of at least 1,250 patients and an equal number of controls to identify a marker with an odds ratio of 2 (assuming a 5% disease prevalence, a 5% minor allele frequency, a power of 80%, and a 5% error rate in each allelic test) (Hong and Park, 2012). Large sample sizes are also needed to study environmental risk factors with low prevalence rates, such as occupational exposures or use of specific medical drugs, and when studying complex gene–environment interactions. To reach adequate numbers, researchers often treat different types of birth defects or childhood cancers as one disorder, whereas their etiology may be vastly different (Wilcox, 2010). Therefore, biobanking and international collaboration seem the only way forward in research on birth defects and childhood cancers.

The AGORA (**A**etiologic research into **G**enetic and **O**ccupational/Environmental **R**isk Factors for **A**nomalies in Children) data- and biobank aims to facilitate research regarding the etiology of major structural birth defects and childhood cancer by routinely collecting data from patients with these disorders and from a population-based control group. These data include DNA samples of children and parents as well as clinical data and parental questionnaires on a wide range of nongenetic risk factors. In this study, we describe the overall design, collection procedures, and data management of the AGORA data- and biobank. In addition, we provide an overview of the patients and controls currently available and give examples of research using these resources.

Materials and Methods

The AGORA data- and biobank was founded at the Radboud university medical center (Radboudumc) in Nijmegen, a referral hospital for many birth defects for approximately 3 million inhabitants living in the east of the Netherlands. AGORA is an ongoing and continuously growing data- and biobank that was initiated by the departments for Health Evidence, Human Genetics, and Urology at the Radboudumc. Data collection was started at the Department of Pediatric Urology in December 2004 and has been extended to other departments or pediatric sub-specialties (surgery, nephrology, cardiology, hemato-oncology, neonatology, human genetics, and the cleft lip/palate and craniofacial center) over the past 10 years. Since 2010, we have also collected data from patients with orofacial clefts at the university hospital in Leuven (Belgium), and in 2013, we started recruiting patients with congenital anomalies of the kidney and urogenital tract (CAKUT) and nephronophthisis at the university medical centers of Utrecht and Amsterdam. Recently, AGORA became part of the Radboud Biobank (www.radboudbiobank.nl), an infrastructure within the Radboudumc for collection, storage, and management of biomaterials and clinical data (Manders et al., 2014). The data collection protocol was approved by the Regional Committee on Research Involving Human Subjects Arnhem-Nijmegen.

STUDY POPULATION: SELECTION AND RECRUITMENT

Birth defects. Since December 2004, parents of children diagnosed with or treated for a birth defect at the Radboudumc have been asked to participate in AGORA. All birth defects are included at the Department of Urology, whereas other departments include specific major birth defects only. Patient recruitment is a routine procedure during the first hospital visit, in which the treating physician provides a short explanation of the aims and procedures of AGORA and hands an information brochure and an invitation letter

to the parents. Informed consent is usually asked for during the second visit to the hospital by a nurse specialist, who has a more independent relationship with the families than the treating physician. When parents (for children <18 years) and patients (for children ≥ 12 years) agree to participate, they sign an informed consent form. Blood of the child is sampled during a surgical procedure or when blood sampling is needed in routine clinical practice. Blood of the parents is sampled while they are in the hospital for the treatment of their child. When blood sampling is not possible, saliva is sampled with a collection kit for use in the hospital or at home. In addition, both parents receive a questionnaire to fill out at home. In rare familial cases only, blood or saliva is sampled from siblings as well.

To rapidly increase the numbers of patients available for research projects on hypospadias, anorectal malformations, CAKUT, and orofacial clefts, we also implemented several retrospective data collection efforts. These efforts included searching medical files for patients treated in the past and inviting them to participate in AGORA, as well as organizing special blood drives for patients and parents. In addition to patients from the Radboudumc, the retrospective data collection efforts comprised patients from medical centers in Leeuwarden, Rotterdam, and Groningen.

Childhood cancer. The inclusion criteria for childhood cancer are: (1) a diagnosis of cancer at least 1.5 years before inclusion (2.5 years for acute leukemia to exclude interference in the genetic results by the disease, which is characterized by altered DNA structure in white blood cells); (2) no chemotherapy for the past 6 months; and (3) no allogeneic bone marrow or stem cell transplantation. The recruitment procedures are similar to those for patients with birth defects. So far, however, the parents have not been asked to fill out the parental questionnaires, as these need to be modified with specifics related to childhood cancer and extended to also cover the time period from birth to diagnosis.

Controls. In the early years of AGORA, we used controls from other studies or performed genetic and gene–environment interaction analyses using case–parent triad designs. In 2010–2011, controls were recruited through random sampling from municipal registries in the Netherlands to better match the AGORA patient population. These municipalities were chosen to cover the referral areas of the hospitals through which the patients were included in AGORA. In total, 42 municipalities were asked to provide a random sample of their inhabitants in the age range of 0 to 20 years. Thirty-nine villages and cities agreed to participate and provided a random sample of 150 or 300 children, respectively. The age distribution of the controls was chosen to be comparable to that of the patient population at time of recruitment. A total of 6,669 families were asked to fill out the AGORA questionnaires, and, due to budgetary restraints, 2,810 families (42%) were asked to also collect saliva samples from child and mother. To update the control group, we may include children from the prospective PRegnancy and Infant DEvelopment (PRIDE) Study in the near future (van Gelder et al., 2013).

CLINICAL DATA

Medical records of the patients are reviewed by medical doctors, clinical geneticists, and researchers to retrieve detailed diagnostic information about specific birth defects or cancer types, and information on possibly known causes, associated malformations, and potential syndromes. The medical records contain information assessed by experienced pediatric medical specialists during physical examination of the child before or during surgery. For some birth defects, detailed clinical information is collected, for example, about the anatomical location of the urethral opening in hypospadias or the specific type of anorectal malformation. For controls, information on the presence of birth defects has been derived from the parental questionnaires.

DNA SAMPLING/ISOLATION

Blood is sampled in ethylenediaminetetraacetic acid tubes, while saliva is collected using Oragene containers (DNA Genotek Inc., Ottawa/Ontario/Canada). DNA is isolated from blood or saliva using standard methods and is stored at -20°C, mainly at the Human Genetics laboratory of the Radboudumc, which was granted accreditation for quality control by the coordinating committee for improvement of quality control of laboratory research in health care.

QUESTIONNAIRE DATA

To obtain information about potential nongenetic risk factors, both parents receive a questionnaire. These questionnaires contain questions about demographic factors, family history of birth defects and cancers, and pregnancy history. In addition, questions about health status, prescribed medication, lifestyle, and various exposures at work or during leisure time activities are included. These questions pertain to the 3 months before conception for fathers and concern both the 3 months before conception and the entire pregnancy for mothers (see Table 1). Mothers are also asked about specific time windows of exposure.

DATA MANAGEMENT

All AGORA data have been stored in an online database. The privacy of participants is guaranteed by an extensive and secure user management application. Clinical information and details about available biomaterials (e.g. DNA from blood or saliva, dates of blood draw and DNA isolation) are manually entered into the database. Questionnaire data are entered either manually or electronically from the reading program Teleform (Cardiff Software Inc., Vista, CA). Items from the questionnaires are cleaned by AGORA researchers when they are used for the first time. The data cleaning procedures are then registered in a data cleaning protocol especially developed for AGORA. This

TABLE 1. *Topics Addressed in the Maternal and Paternal AGORA Questionnaires*

Topics	Maternal	Paternal
Family demographics		
Race/ethnicity	X	X
Education	X	X
Age at childbirth	X	X
Family history of birth defects or cancer in first or second degree relatives	X	X
Health[a]		
Chronic diseases (e.g. diabetes, asthma)	X	X
Pregnancy complications (e.g. preeclampsia)	X	
BMI (pre-pregnancy)	X	X
Bladder, kidney, or urinary tract infection	X	
Other infections	X	
Fever	X	X
Prescribed medications	X	X
Nausea	X	
X-rays	X	
Surgery	X	
Vaccination	X	
Stress	X	X
Pregnancy		
Pregnancy history	X	
Fertility problems	X	X
Assisted reproductive techniques	X	
Time-to-pregnancy	X	
Contraception	X	
Multiple pregnancy	X	
Gestational age	X	
Birth weight	X	
Lifestyle[a]		
Smoking	X	X
Alcohol consumption	X	X
Coffee, tea, or cola consumption	X	X
Folic acid / multivitamin supplements	X	X
Cosmetics	X	X
Diet	X	X
Vegetarianism	X	X
Home[a]		
Living environment (e.g. city, village)	X	
Incineration facility nearby	X	
Use of organic solvents	X	X
Use of pesticides	X	X
Use of other chemicals	X	X
Work[a]		
Occupation	X	X
Use of organic solvents	X	X
Use of pesticides	X	X
Use of X-rays, anesthetics, or cytostatics	X	X
Exposure to exhaust fumes	X	X
Exposure to welding fumes	X	X
Use of other chemicals	X	X

[a]These questions pertain to the 3 months before conception for fathers and concern both the 3 months before conception and the entire pregnancy for mothers, who are also asked about specific time windows of exposure.

protocol will be used every time the same data are needed in a new study.

AGORA COORDINATION AND USE OF DATA AND BIOMATERIALS

AGORA is coordinated by a project group consisting of epidemiologists, geneticists, and clinicians. The project group developed a document with guidelines for the collection and use of data, as well as for collaboration and authorships. Researchers who want to use data and/or DNA are invited to fill out an application form that will be reviewed by the AGORA project group, the medical specialists responsible for the disorder to be studied, and the Radboud Biobank. Once the project is approved, a Material Transfer Agreement (MTA) will be written and signed, after which samples and data will be provided. Application forms for project proposals are available at the AGORA Web site (www.agoraproject.nl).

Several studies have already been performed using data and biomaterials collected within AGORA. In the results section, these studies are briefly described.

Results

Since December 2004, data have been collected from over 6,860 families, including 3,747 families having a child with a birth defect, 905 families having a child with cancer, and 2,208 control families (Table 2).

PHENOTYPIC DESCRIPTION

The types of birth defects among the children included vary widely and comprise malformations of the respiratory, digestive, and urogenital tracts, as well as facial, cardiovascular, kidney, skeletal, and central nervous system anomalies (Table 3). The largest groups of patients include

hypospadias ($n = 944$), anorectal malformations ($n = 563$), orofacial clefts ($n = 468$), and vesico-ureteral reflux (VUR) ($n = 389$). Seventy percent of patients were diagnosed with an isolated defect, ranging from 13% for cardiovascular anomalies to 97% for orofacial clefts. In the remaining 30%, the additional malformations often belong to the same diagnostic category, such as an anorectal malformation occurring together with esophageal atresia, or ureter-pelvic-junction obstruction in combination with renal hypodysplasia. Sixty-four patients (2.2%) presented with the VACTERL association, defined as a combination of at least three anomalies out of the spectrum of Vertebral, Anorectal, Cardiac, Trachea-Esophageal, Renal, and Limb anomalies. The most common childhood cancer types in AGORA comprise acute lymphatic leukemia, Hodgkin and non-Hodgkin lymphoma, Wilms' tumor, and brain and spinal cord tumors (Table 4). Most parents of control children reported that their children were healthy, but 55 controls seemed to have a major birth defect. These families have been excluded from the genetic and nongenetic risk factor analyses for specific anomalies.

AVAILABILITY OF DNA AND QUESTIONNAIRES

DNA is available from 77% of the 6,860 families. These percentages are 96% for families with a child having a birth defect, 100% for families that have a child with cancer (and have not yet been asked to fill out questionnaires), and 34% for control families. The latter percentage is lower by design, as only 42% of control families were asked to provide DNA (Table 2). DNA samples are available from complete case–parent triads in 70% of the families who provided DNA, and from at least the patient and their mother in 80%. DNA was mostly isolated from blood, but saliva was used in 26%.

Questionnaire data are available from 5,049 families (74%). These percentages are 82% for the families with a child having a birth defect, while they are 0% for families that have a child with cancer and 100% for control families by design. In the vast majority of families, both parents filled out the questionnaires, but only the mother did so in 402 families (8%) and only the father in 38 families (1%). In 55 families, questionnaires were filled out for more than one child.

RESPONSE RATES AND DEMOGRAPHIC CHARACTERISTICS

Exact response rates for the prospective data collection cannot be calculated because the number of families that have been invited to participate in AGORA is unknown. For the retrospective data collection efforts, however, response rates were between 65% and 75%. The response rates for the routine AGORA data collection are likely to be higher, due to direct contact with the families and sample collection during regular hospital visits. The response rates among the control group were 34% for the parental questionnaires and 28% for saliva collection. Comparisons between the families with children having birth defects or cancer and control families showed that the percentage of boys was higher in the patient group due to the large numbers of patients with male gender-specific malformations, such as hypospadias and cryptorchidism (Table 5). Only a small percentage of patients (4.9%) and controls (3.8%) were part of a twin. The time lag between childbirth and filling out the questionnaires was 8.9 years on average, but it was 2.5 years shorter for parents of patients compared with control parents. Maternal and paternal ages at childbirth were comparable between patients and controls. The parental educational levels did not differ much between the two groups either, although there seemed to be a slightly higher level of education among control parents. The majority of patient and control parents were of Dutch descent (>91%).

TABLE 2. *Numbers of Questionnaires and DNA Samples from 6,860 AGORA Families Having a Child with a Birth Defect or Childhood Cancer or a Control Child*

	Birth defect	Childhood cancer	Controls
No. of families	3,747	905	2,208
No. of questionnaires			
Mother	2,869[a]	-	2,197[b]
Father	2,727[a]	-	1,973
No. of DNA samples collected	9,667	2,114	1,493
Child	3,482	905	747
Blood	2,679	900	-
Saliva	803	5	747
Mother	3,231	651	746
Blood	2,652	580	-
Saliva	579	71	746
Father	2,954	558	-
Blood	2,355	449	
Saliva	599	109	

[a]In 53 families, questionnaires were filled out for two children in the same family.

[b]In two families, questionnaires were filled out for two children in the same family.

RESEARCH RESULTS FROM STUDIES USING AGORA DATA AND BIOMATERIALS

The first studies using data and biomaterials from the AGORA data- and biobank focused on the etiology of hypospadias and included a replication study of the most important genetic associations reported at the time (van der Zanden et al., 2010), as well as a gene–environment interaction analysis to figure out whether the lack of replication could be due to differences in environmental

TABLE 3. *Types of Birth Defects in AGORA by Diagnostic Category and Phenotype, Based on 3,747 Children with One or More Birth Defects*

Category and phenotype	Number	Category and phenotype	Number
Digestive tract[a]	651	**Orofacial clefts**	468
Anorectal malformations	563	Cleft lip with/without cleft palate	284
Oesophageal atresia	54	Cleft palate only	120
Duodenum/ileum/small intestine atresia	31	Pierre Robin sequence	15
Other	62	Unknown type	49
Abdominal and respiratory tract	54	**Central nervous system anomalies**[a]	41
Diaphragmatic hernia	22	Spina bifida	11
Omphalocele	16	Microcephalia	2
Gastroschisis	9	Tethered cord	12
Respiratory tract anomalies	7	Other	31
Urogenital tract[a]	1,836	**Kidney anomalies**[a]	616
Hypospadias	944	Unilateral renal agenesis	44
Meatus stenosis	51	Renal hypodysplasia	80
Phimosis	55	Ureter-pelvic-junction obstruction	194
Cryptorchidism	177	Hydronephrosis	206
Vesico-ureteral reflux (VUR)	389	Duplex collection system	135
Ureter anomalies[b]	71	Multicystic dysplastic kidney	55
Bladder anomalies	50	Ectopic kidney	13
Posterior urethral valves (PUV)	203	Other[e]	82
Other urologic anomalies[c]	200		
Other genital anomalies[d]	144		
Cardiovascular anomalies[a]	128	**Skeletal anomalies**[a]	117
Ventricular septal defect	51	Vertebral anomalies	73
Atrial septal defect	38	Extremity anomalies	32
Aortic coarctation	9	Rib anomalies	18
Persistent ductus arteriosus	18	Hip dysplasia	21
Other	62	Other	23

[a]The total number of patients with one or more anomalies in a category.

[b]Excluding ureter-pelvic-junction obstruction, but including mega-ureter and ectopic ureter.

[c]Including functional bladder dysfunction, inguinal hernia, and utriculus cysts.

[d]Including vagina and uterus anomalies, hydrocele, and epispadias.

[e]Including enlarged kidney and horseshoe kidney.

exposures among populations (van der Zanden et al., 2012). In a GWAS with pooled DNA samples, we identified the X-chromosomal gene encoding diacylglycerol kinase κ, *DGKK*, as a major risk gene for hypospadias (van der Zanden et al., 2011). Another study, focused on nongenetic risk factors in hypospadias subgroups, provided clear indications for etiologic heterogeneity of hypospadias (van Rooij et al., 2013). The AGORA samples from hypospadias patients have also been used as replication cohorts in studies by other research groups (Geller et al., 2014; Soderhall et al., 2014).

In recent years, our etiologic studies have been focused on anorectal malformations, the VACTERL association, and CAKUT. We found several nongenetic factors that were associated with anorectal malformations, but no association was observed with the use of folic acid supplements (van Rooij et al., 2010; Wijers et al., 2014, 2015). In a small study on patients affected with both an anorectal malformation and hypospadias, we identified a potentially interesting nonsynonymous variant in the *DKK1* gene (van de Putte et al., 2014). Using VACTERL patients from AGORA, mutations in the *TRAP1* gene were found to be

TABLE 4. *Types of Childhood Cancer in AGORA, Based on 905 Children with Childhood Cancer*

Type	Number	Type	Number
Acute lymphatic leukaemia	271	**Brain and spinal cord tumours**	184
Hodgkin Lymphoma	69	Astrocytoma	44
Non-Hodgkin Lymphoma	88	Neuroblastoma	43
Wilms' tumour	91	Craniopharyngioma	16
Osteosarcoma	22	Medulloblastoma	23
Ewing sarcoma	30	Ependymoma	17
Rhabdomyosarcoma	34	Meningioma	8
Hepatoblastoma	10	Choroid plexus tumour	2
Germ cell tumour (non-CNS)	22	CNS germ cell tumour	8
Optical glioma	11	CNS tumour other	23
Other	99		

CNS, central nervous system

highly likely to cause CAKUT or the VACTERL association with CAKUT (Saisawat et al., 2014). In another study on CAKUT, 208 candidate genes were screened in a large group of patients. We only determined a clear genetic cause in approximately 10% of patients, whereas the etiology seemed to be much more complex in the other patients (Nicolaou et al., 2015). Currently, we are also performing studies on obstructive uropathies and orofacial clefts using the AGORA data- and biobank.

Discussion

To study the etiology of multifactorial disorders, large datasets are needed. Due to low prevalence and incidence rates, however, this is quite a challenge for most birth defects and childhood cancers. The AGORA data- and biobank was set up to collect biomaterials and an array of data from sufficient numbers of children with and without these disorders to perform sound etiologic studies.

STRENGTHS AND LIMITATIONS OF THE AGORA DATA- AND BIOBANK

A major strength of the AGORA data- and biobank is that large and unique samples of patients with birth defects and childhood cancers have been collected. The AGORA research team brings together expertise from many different disciplines, including pediatrics, clinical genetics, and epidemiology. Extensive diagnostic information has been gathered for most patients, which allows studying isolated defects separately from nonisolated defects and distinguishing between different phenotypes of certain malformations and childhood cancers. In addition, AGORA includes data from a population-based control group derived from the same geographical area as the patients with a comparable age distribution at time of recruitment, which enables us to make straightforward case–control comparisons.

DNA has been extracted from blood or saliva, both providing high-quality DNA that can be used in all genotyping techniques, including high-throughput techniques, such as exome and genome sequencing (Kidd et al., 2014). The availability of parental samples allows us to determine whether genetic variants are de novo or inherited. In addition, it allows testing for imprinting, the phenomenon by which expression levels of alleles depend upon their parental origin (parent-of-origin effect), and for genetically mediated parental effects. The latter occurs because parental genes are not only transmitted to the child, but also affect parental phenotypes that may in turn influence embryonic development. Furthermore, parental DNA enables the use of family-based association tests that are robust to population stratification. Due to the availability of DNA from healthy control children, traditional case–control analyses can also be performed. These analyses are more efficient in terms of number of samples needed to genotype. The combination of both provides extra power in genetic association studies (Lasky-Su et al., 2010).

The broad range of topics addressed in the parental questionnaires enables explorative as well as hypothesis-testing research into many different nongenetic factors. By combining these data with genetic data, the presence of gene–environment interactions can also be tested.

Due to the rarity of most birth defects and childhood cancers, a prospective study design with inclusion of pregnant women is virtually impossible in terms of time and efficiency. The retrospective design of AGORA is much more realistic, but has its limitations. Retrospective data collection imposes no problems for genetic data, but it may lead to misclassification due to recall problems for questionnaire information. Misclassification only results in recall bias when it is differential, for example, when parents of patients remember or report past events

TABLE 5. *Demographic characteristics of 3,747 children with birth defects and 2,208 control children and their parents.*

Characteristic	Patients with birth defects	Controls
Child	*N=3,747*	*N=2,208*
Gender[a]		
Boy	2,756 (74.3%)	1,082 (49.2%)
Girl	953 (25.7%)	1,118 (50.8%)
Twin[b]	137 (4.9%)	83 (3.8%)
Dizygotic	97 (71%)	54 (65%)
Monozygotic	25 (18%)	16 (19%)
Unknown	15 (11%)	13 (16%)
Mother	*N=2,816*[b,c]	*N=2,195*[b,d]
Time lag between childbirth and filling out the AGORA questionnaire in years - mean (SD)	7.8 (6.7)	10.3 (6.0)
Age at childbirth in years - mean (SD)	31.0 (4.3)	31.4 (4.1)
Educational level[e]		
Low	382 (14%)	221 (10%)
Middle	1,480 (53%)	1,170 (54%)
High	941 (33%)	797 (36%)
Ethnicity		
Dutch	2,470 (92.5%)	1,956 (91.4%)
European other than Dutch	87 (3.3%)	46 (2.2%)
Turkish	28 (1.0%)	13 (0.6%)
Moroccan	10 (0.4%)	5 (0.2%)
Surinamese/Antillean/African	13 (0.5%)	14 (0.7%)
Asian	28 (1.0%)	28 (1.3%)
Other	35 (1.3%)	77 (3.6%)
Father	*N=2,674*[b,c]	*N=1,973*[b]
Time lag between childbirth and filling out the AGORA questionnaire - mean (SD)	7.6 (6.6)	10.1 (6.0)
Age at childbirth in years - mean (SD)	33.6 (4.8)	34.2 (4.9)
Educational level[e]		
Low	506 (19%)	337 (17%)
Middle	1,215 (47%)	864 (44%)
High	891 (34%)	762 (39%)
Ethnicity		
Dutch	2,355 (94.1%)	1,774 (92.3%)
European other than Dutch	55 (2.2%)	34 (1.8%)
Turkish	25 (1.0%)	13 (0.7%)
Moroccan	11 (0.5%)	4 (0.2%)
Surinamese/Antillean/African	12 (0.5%)	16 (0.8%)
Asian	26 (1.0%)	14 (0.7%)
Other	18 (0.7%)	68 (3.5%)

[a]Gender was unknown for 38 children

[b]Information was derived from questionnaire data. Totals for each characteristic may not equal the total number of questionnaires available since some parents did not answer all questions

[c]In 53 families, questionnaires were filled out for 2 children in the same family and data from the oldest child only were used

[d]In 2 families, questionnaires were filled out for 2 children in the same family and data from the oldest child only were used

[e]Educational level was subdivided into: low = no, primary, or lower vocational education; middle = intermediate secondary, intermediate vocational, or higher secondary education; and high = higher vocational or university education.

differently from control parents. Nondifferential forms of misclassification may cause attenuation of the results rather than bias. Our data may be subject to some misclassification or recall bias, as controls have a slightly higher educational level and the time elapsed between childbirth and filling out the questionnaires was shorter for patients (7.7 years) than for controls (10.2 years). To reduce this difference and update our control group, we will include children and their parents from the prospective PRIDE Study (van Gelder et al., 2013) in AGORA in the near future. Because we oversampled controls, however, we are already able to select subgroups of controls that are frequency-matched for date of birth and/or lag-time in studies of specific birth defects. In addition, AGORA also provides the opportunity to use children with other birth defects as controls, which could make sense as parents of affected children may be more likely to report exposures than parents of healthy children (Werler et al., 1989).

With our sampling procedure for childhood cancers, we introduce a selection toward survivors. Therefore, we may miss the most severe cases of cancer.

OPPORTUNITIES FOR RESEARCH AND COLLABORATION

The large numbers of DNA samples, clinical data, and information on a range of nongenetic factors allows for explorative or more detailed etiologic studies on a broad spectrum of birth defects and childhood cancers, as shown by our own studies. Depending on the research question and type of disorder, genetic material can be used for a wide range of analyses, such as candidate gene approaches, copy number variation analyses, GWASes, and exome- or whole-genome sequencing. For several birth defects and childhood cancers, the numbers of samples collected are sufficient to

provide a discovery cohort, especially because very few genetic studies have been performed for these disorders, leaving the low-hanging fruit yet to be discovered (van der Zanden et al., 2011). For other disorders with smaller samples sizes in AGORA, data and biomaterials may still be used in hypothesis-validating studies, such as genetic replication studies, or in larger studies with a combination of cohorts from several research groups. The available information on nongenetic factors may provide a valuable addition to genetic studies or could be used in epidemiologic analyses. As international collaboration is the key to success for studies on multifactorial disorders, AGORA provides a large and unique set of data and biomaterials that is available for external researchers. This will provide a collective opportunity to add to the knowledge base for birth defects and childhood cancers and to identify more potentially causal genetic and nongenetic factors and their interactions, which may eventually lead to improved diagnosis, genetic counseling, and prevention of these disorders.

Acknowledgements

We are grateful to all staff members involved at the departments of Urology, Surgery, Pediatrics, Genetics, Orthodontics, and Craniofacial Biology, and the cleft lip and palate and craniofacial center of the Radboudumc, with special thanks to the specialized nurses Karen Kwak, Jacqueline Knoll, Mariette van de Vorle, Marlaine Hammen, Marjan Kloen, Marietje van Mullekom, Rita Scholtens, Hans Coolen, Annelien Schulp (UMCU), and research assistant Ursula Oldenhof for their collaboration in collecting informed consent approval and data of patients and controls. We thank Saskia van der Velde-Visser, Mieke Kipping-Geertsema, and the administrative team of DNA diagnostics of the Department of Human Genetics for their help in registering the DNA samples and the group of Barbara Franke for the DNA isolation from the saliva samples, with special thanks to Johanne Groothuismink. We acknowledge Ariaan Siezen for her assistance in building the AGORA database. We are also grateful to staff members involved at the departments of Medical Genetics and Pediatric Nephrology of the University Medical Center Utrecht, The Netherlands, and the Department of Oral Health Sciences-Orthodontics and the orofacial cleft team, KU Leuven, University Hospital of KU Leuven, Belgium.

References

Buka I, Koranteng S, Osornio Vargas AR. 2007. Trends in childhood cancer incidence: review of environmental linkages. Pediatr Clin North Am 54:177–203.

Corsello G, Giuffre M. 2012. Congenital malformations. J Matern Fetal Neonatal Med 25;(Suppl 1):25–29.

Geller F, Feenstra B, Carstensen L, et al. 2014. Genome-wide association analyses identify variants in developmental genes associated with hypospadias. Nat Genet 46:957–963.

Hong EP, Park JW. 2012. Sample size and statistical power calculation in genetic association studies. Genomics Inform 10:117–122.

Kaatsch P. 2010. Epidemiology of childhood cancer. Cancer Treat Rev 36:277–285.

Kidd JM, Sharpton TJ, Bobo D, et al. 2014. Exome capture from saliva produces high quality genomic and metagenomic data. BMC Genomics 15:262.

Lasky-Su J, Won S, Mick E, et al. 2010. On genome-wide association studies for family-based designs: an integrative analysis approach combining ascertained family samples with unselected controls. Am J Hum Genet 86:573–580.

Manders P, Siezen AE, Gazzoli S, et al. 2014. Radboud Biobank: a central facility for prospective clinical biobanking in the Radboud university medical center, Nijmegen. OA Epidemiology 10:1–4.

Nicolaou N, Pulit SL, Nijman IJ, et al. 2015. Prioritization and burden analysis of rare variants in 208 candidate genes suggest they do not play a major role in CAKUT. Kidney Int [Epub ahead of print].

Saisawat P, Kohl S, Hilger AC, et al. 2014. Fine mapping analysis confirms and strengthens linkage of four chromosomal regions in familial hypospadias. Eur J Hum Genet 23:516–522.

Soderhall C, Korberg IB, Thai HT, Cao J, Chen Y, Zhang X, Shulu Z, van der Zanden LF, van Rooij IA, Frisen L, Roeleveld N, Markljung E, Kockum I, Nordenskjold A. 2014. Fine mapping analysis confirms and strengthens linkage of four chromosomal regions in familial hypospadias. Eur J Hum Genet 23:516–522.

Spector LG. 2010. Assessing parental contributions to childhood cancer risk. Future Oncol 6:5–7.

Spector LG, Hooten AJ, Ross JA. 2008. Ontogeny of gene expression: a changing environment for malignancy. Cancer Epidemiol Biomarkers Prev 17:1021–1023.

Stiller CA. 2004. Epidemiology and genetics of childhood cancer. Oncogene 23:6429–6444.

van de Putte R, Wijers CH, de Blaauw I, et al. 2014. Sequencing of the DKK1 gene in patients with anorectal malformations and hypospadias. Eur J Pediatr 174:583–587.

van der Zanden LF, Galesloot TE, Feitz WF, et al. 2012. Exploration of gene-environment interactions, maternal effects and parent of origin effects in the etiology of hypospadias. J Urol 188:2354–2360.

van der Zanden LF, van Rooij IA, Feitz WF, et al. 2011. Common variants in DGKK are strongly associated with risk of hypospadias. Nat Genet 43:48–50.

van der Zanden LF, van Rooij IA, Feitz WF, et al. 2010. Genetics of hypospadias: are single-nucleotide polymorphisms in SRD5A2, ESR1, ESR2, and ATF3 really associated with the malformation? J Clin Endocrinol Metab 95:2384–2390.

van Gelder MM, Bretveld RW, Roukema J, et al. 2013. Rationale and design of the PRegnancy and Infant DEvelopment (PRIDE) Study. Paediatr Perinat Epidemiol 27:34–43.

van Rooij IA, van der Zanden LF, Brouwers MM, et al. 2013. Risk factors for different phenotypes of hypospadias: results from a Dutch case-control study. BJU Int 112:121–128.

van Rooij IA, Wijers CH, Rieu PN, et al. 2010. Maternal and paternal risk factors for anorectal malformations: a Dutch case-control study. Birth Defects Res A Clin Mol Teratol 88: 152–158.

Werler MM, Pober BR, Nelson K, Holmes LB. 1989. Reporting accuracy among mothers of malformed and nonmalformed infants. Am J Epidemiol 129:415–421.

Wijers CH, de Blaauw I, Zwink N, et al. 2014. No major role for periconceptional folic acid use and its interaction with the MTHFR C677T polymorphism in the etiology of congenital anorectal malformations. Birth Defects Res A Clin Mol Teratol 100:483–492.

Wijers CH, van Rooij IA, Rassouli R, et al. 2015. Parental subfertility, fertility treatment, and the risk of congenital anorectal malformations. Epidemiology 26:169–176.

Wilcox AJ. 2010. Birth defects. In: Wilcox AJ, editor. Fertility and pregnancy: an epidemiologic perspective. New York: Oxford University Press. pp. 230–245.