COMING OF AGE

Researchers in Britain have tracked thousands of children since their birth in the 1990s. Now the study is 21, and turning to the next generation.

BY HELEN PEARSON

ALSPAC (CHILDREN OF THE 90S)

Amy Murdoch-Davis is one of 14,000 people who have been studied since birth in and around Bristol, UK; her baby, Esmé, will soon join the study, too.

In a secure storage barn on the outskirts of Bristol, nearly 9,000 placentas float in plastic buckets of formaldehyde.

Twenty-five kilometres away, in the basement of a university building, baby teeth from more than 4,000 children fill cardboard boxes in a walk-in freezer. Next door are some 15,000 nail clippings and 20,000 locks of hair. A few steps farther, a parade of freezers house row upon row of bar-coded blood cells, plasma, urine, saliva and chunks of umbilical cord that together make up a tissue library with more than one million entries.

This library is the harvest from an unusual study of humanity. In 1990, researchers started to collect tissues and detailed information from more than 14,500 pregnant women in this western British city and its surrounding region of Avon. The women filled in more than 100 pages of questionnaires about their health, relationships, work and home. After birth, researchers tracked the children's development through surveys, clinical examinations and biological samples. They know what the kids ate, when they first talked, how often they fell sick and when a parent read to them — or deserted them. They know when the children started to hit puberty, drink alcohol, have sex and leave home. In that wealth of data — collected at a cost of some £42 million (US$67 million) so far — they are tracing how genetic and environmental factors in the children's early years affect their later ones.

Now, the Bristol cohort study is coming of age, literally and scientifically. This spring, the first members of the group turn 21. And on 18 April, leaders of the study and their collaborators from around the world will meet in Bristol to discuss what they have learned from the Avon Longitudinal Study of Parents and Children (ALSPAC), also known as the Children of the 90s. The rich collection of tissues and behavioural data — a comprehensive phenotype for each of thousands of participants — makes the study unique and invaluable, say its leaders. "It's the deepest phenotyping and biobank resource of any large birth cohort — unequivocally," says George Davey Smith, the study's scientific director. The results have generated more than 700 scientific papers and range from policy-changing health advice for pregnant women and young children to the discovery of genetic factors involved in fetal growth, obesity, allergies and bone density[1–4]. The study has also inspired and guided other birth cohorts, including the world's largest, which is tracking more than 100,000 children in Norway.

Yet many of the data and samples remain untouched, and the cohort leaders acknowledge that the most important findings are likely to emerge years from now, some of them courtesy of new techniques that the founders of the study could barely have imagined. Davey Smith and his colleagues are just starting to analyse the children's genomes — 2,000 of which have been fully sequenced — and epigenomes, the catalogue of chemical footprints left on a child's DNA by experiences in the womb or in the world. The researchers say that such studies will help them to move from a slew of loose epidemiological associations — between a mother's fish-eating and her child's intelligence, for example — to the genetic and epigenetic players responsible for those links.

As the children themselves become parents, the team is expanding the scope of the study to a new generation. Still, at the heart of all this remains an unanswered question: how much value is there in collecting a huge amount of

ALSPAC (CHILDREN OF THE 90S)

The Bristol study has banked more than 1 million samples including blood, urine, saliva and placentas.

information about human life when the scientific pay-off is unknown? Marcus Pembrey, who was director of genetics at ALSPAC until 2005, recalls that when he initially told colleagues that he wanted to measure "everything" about the children, "they laughed and said you can't study everything. And I said you have to study everything. You don't know what will be important in the future."

CONCEPTION AND BIRTH

Today, the effort to gather everything is headquartered in a concrete slab of a building in central Bristol. On the ground floor, a brightly painted annex houses a clinic where the children of the 90s come for their regular medical examinations. A few middle-aged men — the dads of the cohort members — are here, going from room to room for blood-taking, cognitive testing and bone scans. The study scientists are bringing in as many dads as they can to collect clinical data such as DNA, height, weight and blood pressure, in an effort to ramp up studies on the fathers' health, as they already have for the mothers.

Also in the clinic is 19-year-old cohort member Amy Murdoch-Davis, with her five-month-old baby Esmé. Last month, ALSPAC received provisional ethics approval to start recruiting all the children of its cohort members, an effort it hopes to launch in earnest later this year. Esmé will be one of the first signed up. "I've been a guinea pig all my life," Murdoch-Davis says, "and she can be a guinea pig too." This time around, the researchers want to collect even more samples, including breast milk and the babies' first stools.

The study itself was a struggling infant when Murdoch-Davis was born. Its leader, Jean Golding, a mathematician turned epidemiologist, had studied stillbirths and neonatal deaths and was eager to learn how events in pregnancy and infancy affect a child's health and development. She was convinced that studying a large cohort, starting in pregnancy and gathering as much information as possible, was the best way to find answers. Funders, however, saw it differently. "Everybody thought 'if you have a cohort study you're tied down to carrying on funding it — and it's a bottomless pit'," she says.

Golding eventually scraped together some money by writing grant applications that focused on specific diseases. She drummed up interest from mothers by talking on television and radio, and sending an army of midwives to antenatal classes and doctor's surgeries. The 14,541 pregnancies eventually included in the study encompassed more than 70% of those eligible in the region between April 1991 and December 1992. The buckets of fresh placentas started stacking up. And so did other data, such as blood squeezed from the babies' heels during the early clinic visits and reams of questionnaires filled out by the mothers (see 'Building a bank of life'). Little was left unasked. Does baby drink breast milk or formula? Has he or she had antibiotics or skin ointment? Do you have a telephone; a tumble dryer; cockroaches?

Within a year or two of starting, however, the cash was draining away. Golding employed her 40–50 staff one month at a time, never sure whether she would have enough to pay them again. And there was no time or money to analyse the data that were flooding in. "We were well in the red," Golding says. "I was just exhausted."

> "IT'S THE DEEPEST PHENOTYPING AND BIOBANK RESOURCE OF ANY LARGE BIRTH COHORT."

CHILDHOOD AND ADOLESCENCE

A few years later, just as parents were collecting baby teeth from beneath thousands of pillows and mailing them to the research team, the first data analyses started coming through. Researchers now point to a handful of results for which the cohort is famous, and that had an impact on public-health policy. One showed that eating oily fish during pregnancy was associated with better eye and cognitive development in children[5,6]. Another helped to cement advice that babies should be put to sleep on their backs to reduce the risk of cot death, by showing that this sleeping position did not cause any developmental delays[7]. A third showed the first association between peanut allergy — an emerging epidemic in Western countries — and peanut oil in baby lotions[8]. Manufacturers soon started identifying the ingredient on labels.

As the children entered their second decade, the study found itself on firm financial ground for the first time. The Wellcome Trust and the Medical Research Council awarded it core funding that has totalled some £21 million for 2001–14, and investigator-led grants have brought in much more.

With the study maturing comfortably, Golding retired in 2005. ("They stopped paying me," as she puts it, still firmly ensconced in her office at the University of Bristol.) Lynn Molloy, a social scientist, took over the managerial reins, and Davey Smith, an epidemiologist passionate about genetics, grasped the scientific ones.

The study entered a productive scientific adolescence even as the cohort members suffered their own teen agonies. Questionnaire data revealed that 19% of 16–17-year-olds cut or otherwise hurt themselves. By age 18, some 5–10% had experienced some form of psychosis — "more common than people had previously suspected", says Stanley Zammit, a psychiatrist working with cohort data at Britain's Cardiff University — even though few are likely to go on to develop schizophrenia or a related condition.

At the same time, genetics was finally entering the picture. In the early years of the project, examining the children's genotypes was just a fond hope. Scientists had announced a draft human genome in 2001, but that was a multibillion-dollar megaproject. Quite what to do with several thousand raw human genomes,

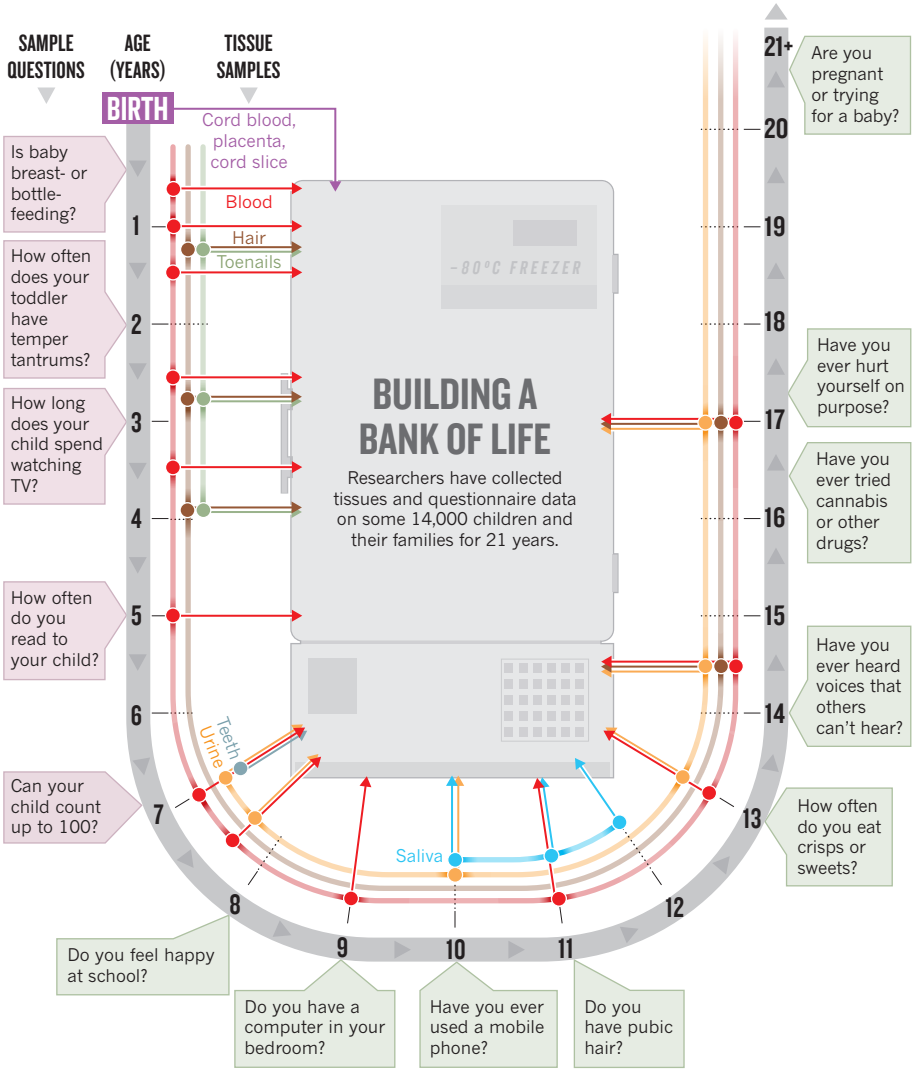
no one knew. "The interesting thing was that although all these high powered alpha-male geneticists were talking the big talk about genetics — when you actually gave them DNA for a thousand individuals it was too much. They didn't have the technology," Golding says.

In 2007, however, data from ALSPAC and several other large human biobanks were used to scour the human genome for single-letter variants associated with obesity. The work turned up a gene called *FTO*, and found that adults with two 'risk' copies of this gene are about 3 kilograms heavier, on average, than those with no risk copies[2]. The discovery became a poster child for the identification of risk alleles through genome-wide association studies, a technique that was sweeping through the genetics community at the time. The Bristol cohort — with its extensive bank of DNA and medical data — was perfectly placed to catch the wave. Its data have been used to identify genetic sequences associated with fetal growth[1], bone density and childhood growth[4], tooth development[9], facial features[10] and more. Researchers are now hunting for genetic links to intelligence, educational attainment and gender orientation.

Davey Smith sees other opportunities to explore associations in the cohort's DNA. He finds it frustrating that epidemiological studies often reveal a correlation between two factors — poverty and obesity, say — without proving that one causes another. The trouble is, many social and biological factors tend to correlate anyway: people who smoke also tend to drink more alcohol, eat unhealthy food, be poorer, weigh more and have high cholesterol and other signature biomarkers of ill health.

One way to filter true causes from the correlations is to compare one cohort with another, something Davey Smith did recently to find 'causal associations' with breastfeeding. In the Bristol cohort, breastfeeding is correlated with less obesity, lower blood pressure, higher intelligence and more good things besides. But in the United Kingdom, breastfeeding mothers are also more likely to be middle- or upper-class. So is it breastfeeding that helps children, or some other aspect of a comfortable life? When Davey Smith and his colleagues looked at a birth cohort based in Pelotas, Brazil, where breastfeeding is not linked to socioeconomic status, the links with obesity and blood pressure fell apart, but the link with IQ held up[11].

Researchers are also finding true causes through Mendelian randomization, a genetic technique that is exciting epidemiologists. Stephanie von Hinke Kessler Scholder, an economist at the University of York, UK, is using this method to unpick the associations between children's obesity and poorer performance in school. Obesity is also correlated with lower socioeconomic status, so Scholder sought to test whether obesity directly hinders performance (because of bullying, for example), whether kids who are obese do less well because they come from disadvantaged families, or whether some other factor could explain the correlation.



Scholder took the banked DNA of the ALSPAC children and divided them up into groups on the basis of the make-up of two weight-related genes, *FTO* and *MC4R*. (Children with all 'heavy' copies of these genes weigh a few kilograms more on average than those with all 'light' copies.) Gregor Mendel's laws of inheritance ensure that the heavy and light forms of the genes are shuffled and randomly delivered to children across a population, irrespective of their social class or any other confounding factor. And when Scholder removed confounders from the equation in this way, she found that children with the heavy versions of *FTO* and *MC4R* did just as well on school tests at age 14 as those who didn't have them. Dispelling the false idea that fat children do worse at school is "a positive thing", she says[12].

The most technologically advanced answer to the causation problem sits upstairs in the Bristol headquarters of the study. There, a £300,000 machine is poised to start rapid analyses of methylation — an epigenetic mark that controls gene activity. By looking at 450,000 sites in the cohort members' genomes, the researchers will build up a bank of methylation data on blood samples taken from the children at birth, at ages 7 and around 17, and on the mothers during pregnancy and 17 years later. "It's a unique resource," says Davey Smith.

In a study published last month[13], Caroline Relton at Newcastle University, UK, who is leading the epigenetic work for ALSPAC, looked at methylation on an array of genes in the umbilical-cord blood of cohort children. She found that methylation signatures on nine genes at birth showed some association with body height and tissue composition at age nine. The finding hints that events during pregnancy might shape gene expression early in life — eventually resulting in altered growth and weight gain. It also "provides a flavour of the sort of thing we can address more powerfully when we get this huge data set under way", Relton says. Next, the team plans to search for methylation patterns linked to factors such as neural development, behavioural problems, cardiovascular health and fatty liver disease. The researchers will

WELLCOME LIBRARY, LONDON

Jean Golding, pictured here with cohort children in 2001, started the study in the early 1990s.

also look for potential causes of these methylation patterns — associations with a mother's smoking, alcohol intake, weight gain during pregnancy, folate levels and exposure to air pollution.

NEXT GENERATION

The methylation robot is just one piece of the high-tech equipment involved in the study. In one lab in Bristol, blood cells taken from the dads that day are being spun down and divided up into aliquots, ready for storage. In another, researchers are transforming cell samples into immortal cell lines. They have already banked around 7,000 such lines, which provide an endless supply of DNA and might be used for future studies of cell behaviour. Sue Ring, the head of the laboratories, says that her team take turns to carry an emergency mobile phone that will warn them of a freezer failure. After all that the participants have given to the study, she says, she feels a "duty of care for the samples".

Soon, the labs will start to process a whole new set of tissues — from the children of the cohort members. The eggs that will develop into these children were formed when their mothers themselves were babies, growing in their own mother's uterus. This means that events in the grandmothers' lives, such as an infection, stress or exposure to toxic chemicals — all recorded in the cohort database — could produce signals in the streams of data that the researchers plan to collect about the grandchildren. "I was going to retire and wind down until we stumbled on these transgenerational events," says Pembrey, who wants to explore the third-generation consequences of traumatic events in the lives of grandparents.

With so much yet to do, has the study justified the years of effort put in so far? Cohort studies sometimes draw criticism because "they are very big and very long-term and they seem to be a lifelong source of income for investigators", says Teri Manolio, who has worked on cohorts in the past and is now director of the Office of Population Genomics at the National Human Genome Research Institute at Bethesda, Maryland.

> "YOU HAVE TO STUDY EVERYTHING. YOU DON'T KNOW WHAT WILL BE IMPORTANT IN THE FUTURE."

But ALSPAC receives kudos from many researchers, including Nigel Paneth, an epidemiologist at Michigan State University in East Lansing, who runs a data-collection site for the US National Children's Study, which started recruiting in 2009. Unlike ALSPAC, the National Children's Study has ballooned into a mammoth programme with costs estimated at as much as $6 billion, and has struggled for more than a decade to convince sceptical scientists and funders that it will pay off scientifically. Paneth says that big cohort studies can make major contributions to health if they are well planned and executed. "Can I guarantee results? Of course I can't. But I can guarantee you no results if we don't do it."

In Bristol, the team is only starting to learn the value of some of the data collected decades ago. The placentas, which had to be moved twice, were regarded mostly as a "bloody nuisance" until three years ago, when David Barker at the University of Southampton, UK, and Oregon Health and Science University in Portland, asked to use them. Barker is famous for drawing connections between early fetal development and adult health; in the past few years he has reported that the size and shape of a placenta — the life support of the fetus — is associated with risk of adult coronary heart disease, high blood pressure and even lifespan[14]. He sent a photographer to Bristol to snap the placentas, and is now using the organs' dimensions to test some of his hypotheses. The correlations are "definitely there and amazingly strong", Barker says.

Other correlations will become clear only when the study and its participants have grown up a little more. They do, after all, have most of their lives ahead of them. Murdoch-Davis is planning the next step in hers: going to university to study English and psychology. Meanwhile, Molloy is mapping out what data to collect when the cohort members next visit the clinic, aged 24–25, when they will be close to the peak of their health. And Davey Smith has faith in future generations of scientists to find new ways to study the cohort. "My hope is that when I drop dead it's got an energetic scientific director who's implementing all these future technologies that I couldn't imagine," he says.

Perhaps, Davey Smith says, someone will be analysing data from the digital video recorders that he's thinking about handing to each expectant family, to document every significant event in their child's life. After all, storing terabytes of data doesn't cost anything — and you never know what they might one day reveal. "Digital recording allows you to store up huge banks of data for future use," he says. "It's a bit like Jean collecting the placentas. No one knew what they were going to do with them." ■

Helen Pearson *is* Nature*'s chief features editor.*

1. Freathy, R. M. *et al. Nature Genet.* **42,** 430–435 (2010).
2. Frayling, T. M. *et al. Science* **316,** 889–894 (2007).
3. Henderson, J. *et al. J. Allergy Clin. Immunol.* **121,** 872–877 (2008).
4. Timpson, N. J. *et al. Hum. Mol. Genet.* **18,** 1510–1517 (2009).
5. Williams, C. *et al. Am. J. Clin. Nutr.* **73,** 316–322 (2001).
6. Daniels, J. L. *et al. Epidemiology* **15,** 394–402 (2004).
7. Hunt, L., Fleming, P. & Golding, J. *Pediatrics* **100,** E11 (1997).
8. Lack, G., Fox, D., Northstock, K. & Golding, J. *N. Engl. J. Med.* **348,** 977–985 (2003).
9. Pillas, D. *et al. PLoS Genet.* **6,** e1000856 (2010).
10. Paternoster, L. *et al. Am. J. Hum. Genet.* **90,** 478–485 (2012).
11. Brion, M.-J. A. *et al. Int. J. Epidemiol.* **40,** 670–680 (2011).
13. von Hinke Kessler Scholder, S. *et al.* Preprint available at www.bristol.ac.uk/cmpo/publications/papers/2011/wp274.pdf
13. Relton, C. L. *et al. PLoS ONE* **7,** e31821 (2012).
14. Barker, D. J. *et al.* In *The Placenta and Human Developmental Programming* (eds Burton, G. J. *et al.*) Ch 2, 5–12 (Cambridge Univ. Press, 2010).