learn from that? When science meets Hollywood, science loses.

That's interesting. Care to elaborate? Part of the problem in trying to make a movie about real animals is that they aren't actors. You can write a script — and there was one — but the chimps didn't always follow it! So, like a good field worker, one had to adapt as the film footage came in and the story line changed. In this day and age of global warming, disease pandemics, and the like, it's essential for biologists to reach out to the general public and communicate what we are learning. What better way to do this than via the silver screen? So when the producers approached us saying that they wanted to make a film about chimpanzees that would adhere to the science and what we know about them, it was a no-brainer to sign on. In retrospect, I was a bit naïve to believe this. In biology, we learn that there are always trade-offs in life. There were cases where telling a good story that will sell in theaters clashed with our scientific understanding of chimpanzees. In these situations, artistic license was favored. At the end of the day, I understand and can live with the decisions that were made, as science doesn't always make for great entertainment.

What is the biggest challenge in your field? As I look back on my career, I realize how lucky I have been. Back in the late 1970s when I began my research on apes, the field was wide open. Scant work had been conducted, funds to conduct field work were flowing, and populations of apes were seemingly everywhere to investigate. All of this has changed. Several long-term field studies of apes have been carried out and continue to this day. We live in a molecular and biomedical age, where an increasingly large part of the funding pot goes to things other than studies of animal behavior and field research. Primates today are endangered everywhere. Habitat loss, hunting, and recurrent outbreaks of infectious disease have decimated large populations of primates across the globe. Sadly, the biggest challenge is to ensure that there will be something to study in the future.

University of Michigan, Department of Anthropology, 1085 South University Avenue, Ann Arbor, Michigan 48109-1107, USA.
E-mail: mitani@umich.edu

Quick guide

GWAS

Jonathan Flint

G-what? GWAS stands for 'genome wide association study', the favoured method for finding genetic variants that increase disease risk. In a GWAS, allele frequencies of common genetic variants are compared between cases (those affected by disease) and controls. Common variants are those with a minor allele frequency greater than about 5% (the frequency varies in different populations). GWASs are also used to find genetic variants that contribute to variation in quantitative traits that are not diseases, such as height or weight. For example, a GWAS of obesity looks for any variant whose genotypes are associated with different mean weights.

How do I carry out a GWAS? Use the following simple four-step recipe: first, collect phenotypic information from thousands of individuals; second, extract DNA and genotype at least 500,000 single nucleotide polymorphisms (SNPs); third, call genotypes and detect association using one of a variety of (relatively) user-friendly software packages; fourth, sift through the results and identify at least one association signal with a *P*-value less than $P < 10^{-8}$ (Figure 1). Success is guaranteed if you work on a disease no one has published on before, or if you can find additional loci in an important disease, such as type 2 diabetes, obesity or Crohn's disease. For the latter, carry out your analysis in a novel ethnic group (East Asia is currently a favourite), or simply double the sample size of the last GWAS to increase power and thereby identify novel loci. Please note though that acronymed consortia — not people — write GWAS papers. GWAS authorship is turning into a field of study in its own right. One report counted 21,007 authorships for 604 GWAS. This is because the sample size needed for GWAS is so huge now, up to a quarter of a million people.

Why do I need such a large sample size? The reason for the large sample is that each genetic locus makes such a small contribution to disease susceptibility. The effect size is usually expressed as an odds ratio — if an odds ratio is two, the increase in risk is twofold. Typical GWAS odds ratios are about 1.1–1.2. For quantitative traits, such as height or weight, the size of the effect is usually expressed as a percentage of the phenotypic variance attributable to the locus. For example, if half the variability in height in a population is due to alleles at one locus, then the locus' effect size would be 50% Typical values are about 0.05%. While the sample size required to detect loci varies from phenotype to phenotype, it is always in the thousands. For Crohn's disease, 2,000 cases were sufficient to identify nine loci. For hypertension, 29,000 individuals were needed to detect ten loci.

And why such a low P-value? If you are used to working with $P < 0.05$ to get your paper published, $P < 10^{-8}$ does seem a little over the top. It's the consequence of testing hundreds of thousands of alternative hypotheses (hundreds of thousands of markers) and this is one reason why you'll need a big sample. Actually, it's not as bad as it sounds, requiring only about an eightfold increase in sample size compared to what you need for the 0.05 level.

Why does GWAS work? The idea behind GWAS is that interrogating variation at a few hundred thousand positions is sufficient to capture the bulk of genetic variation. A remarkably small amount of sequence (relative to our genome's size of three gigabases) is sufficient, because our genomes have a relatively simple haplotypic structure, such that variants in close proximity are highly correlated, forming haplotype blocks. And if you are wondering why we have this particular haplotype structure, then the answer is because of human history: it is due to the exit of our ancestors from Africa about 100,000 years ago imposing a population bottleneck, and the subsequent enormous population expansion.

When doesn't GWAS work? Not all human populations have the same haplotype structure and in some cases this can frustrate GWAS success. For example, in Africa, haplotype blocks are on average smaller, so many more markers are needed to capture the majority of the population's common genetic variation. Standard GWAS approaches don't work so well there. A corollary of the recent human

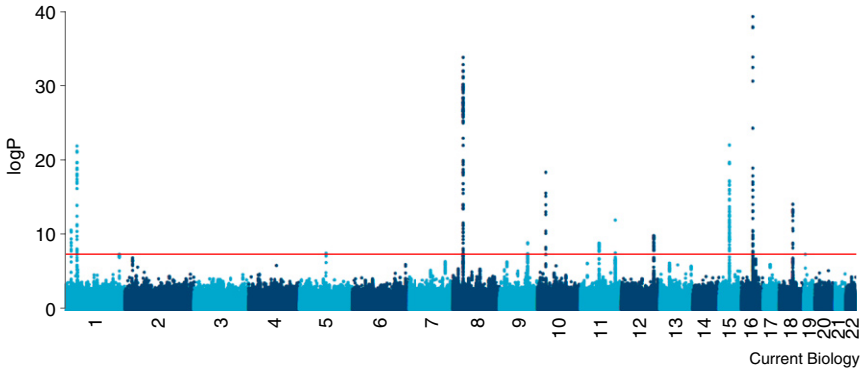

Figure 1. Visualising a GWAS with a Manhattan plot.
The horizontal axis shows the position of every locus that the microarray interrogates. The numbers denote chromosomes. The vertical axis is the negative logarithm (base 10) of the *P*-value (logP) of the association between phenotype and genotype. This plot is based on simulated data, but the experimenters would be pleased, as there are a number of peaks that exceed a genome-wide significance threshold (red line).

population expansion is that most alleles are rare, and are not interrogated by standard commercially available microarrays. The full extent of what is missed became apparent from recent population scale re-sequencing projects: only 13% of variants with a frequency of less than 0.5% had been described previously. If rare variants make a substantial contribution to your disease of interest, beware! GWAS won't find them. You may also have read that GWAS doesn't work because GWAS loci cannot account for much of the known or estimated heritability of a trait ('missing heritability'). For instance, despite finding 180 loci that influence height, these loci account for just 10% of the variation. But, this does not take account of all those SNPs that don't make the significance threshold. They can't simply be ignored, but what to do with them? Peter Visscher has an answer, using an approach routine in plant and animal genetics. Examining the effect of all SNPs, regardless of statistical significance, almost half of height's phenotypic variance can be explained by common SNPs. So is there a 'missing heritability' problem? Well, we still can't explain *all* the variance.

What have we learnt from GWAS?
Two common complaints are that GWAS gives us genetic loci not genes (true!) and that lists of genetic loci don't tell us anything about mechanism (true too!). One of the insights of the ENCODE project is that GWAS hits lie preferentially in regulatory regions of the genome (enhancers, promoters and other less well categorized elements). Tying variation at an enhancer to a particular gene product is admittedly hard, but the nearest neighbouring gene hypothesis works well (ENCODE again helps here, revealing that action on the megabase scale is rare, most elements operate over a few tens of kilobases). Next generation GWAS are now including tests of function, testing gene expression patterns of nearest-neighbour genes in relevant tissues, and (impressively) in a GWAS for human red blood cell phenotypes, haemocyte-specific RNA interference (RNAi) silencing in *Drosophila melanogaster*.

Does this mean GWAS can deliver the holy grail of mechanism? Take note, journal editors, genetics is a hypothesis-free enterprise! How else could mathematicians, statisticians and bioinformaticians partake?

Where can I find out more?

Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D. M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., *et al*. (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.
http://www.nature.com/nrg/series/gwas/index.html
Lander, E.S. (1996). The new genomics: global views of biology. Science *274*, 536–539.
Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R., *et al*. (2001). Linkage disequilibrium in the human genome. Nature *411*, 199–204.
Wellcome-Trust-Case-Control-Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.
Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., *et al*. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. *43*, 519–525.

Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK.
E-mail: jf@well.ox.ac.uk

Correspondences

Children with autism do not overimitate

L. Marsh, A. Pearson, D. Ropar, and A. Hamilton

Copying the behaviour of others is important for forming social bonds with other people and for learning about the world [1]. After seeing an actor demonstrate actions on a novel object, typically developing (TD) children faithfully copy both necessary and visibly unnecessary actions [2]. This 'overimitation' is commonly described in terms of learning about the object, but may also reflect a social process such as the child's motivation to affiliate with the demonstrator [3] or to conform to perceived norms [4]. Previous studies of overimitation do not separate object learning and social imitation because they use novel objects. Even though researchers consider these objects to be causally transparent in their mechanism, young children's causal reasoning about novel objects is unclear [4]. The present study measures the social component of overimitation by using familiar objects, which preclude the learning component of the task. Here we report a significant reduction in overimitation in children with autism spectrum conditions (ASC). This is coherent with reports that these children have profound difficulties with social engagement [5] and do not spontaneously imitate action style [6] (see also [7]).

We tested 31 children with ASC, 30 TD children matched for verbal mental age and 30 TD children matched for chronological age on an overimitation task using familiar objects. All children were assessed for verbal mental age, overimitation and understanding of action rationality (see Supplemental Information). On each of five trials, the child was asked to watch carefully as a demonstrator showed how to retrieve a toy from a box or build a simple object. Critically, each demonstration included two necessary actions (such as unclipping and removing the box lid) and one unnecessary action (such as