

HOW MY GENOME WAS HACKED

If a *New Scientist* reporter's DNA is vulnerable, so is yours. **Peter Aldhous** and **Michael Reilly** investigate

MARK RICHARDS FOR NEWSCIENTIST

INTIMATE secrets hidden in your DNA could be stolen without you even realising. By taking a glass from which you have drunk, a “genome hacker” could obtain a comprehensive scan of your genome, revealing DNA variants that help determine your susceptibility to a wide range of diseases, from a common form of blindness to Alzheimer’s disease.

That’s the disturbing finding of a *New Scientist* investigation, in which one of us – Michael Reilly – “hacked” the genome of the other – Peter Aldhous – armed with only a credit card, a private email account and a home address.

You might have thought that genome hacking requires specialist skills, and personal access to sophisticated equipment. But in recent years, some companies have started to offer personal genome scans to the public over the internet. Other firms routinely analyse genomes on behalf of scientists involved in human genetics research. In theory, both types of service are vulnerable to abuse by a genome hacker determined to submit someone else’s DNA for covert analysis.

Until our investigation, it was not clear whether this would be possible in practice. Could a hacker with no access to a genetics lab take an item carrying another person’s DNA and obtain a sample that companies would accept for scanning? Would the sample be of

“We should have a right to expect that people can’t snoop through our genes”

high enough quality to yield accurate results? And would genome analysis companies have procedures in place to identify and refuse suspicious orders?

We decided to find out. Rather like computer security researchers who expose vulnerabilities in software code so that they can be “patched” to guard against malicious hackers, our goal was to uncover vulnerabilities in the way companies offering genome scans operate, so that they can be fixed.

Our investigation uncovered some loopholes that might be closed to help thwart genome thieves. The findings also strengthen the case for additional laws to protect the information contained in the DNA that we all shed continually and leave lying around.

“Just as we have a right to expect that relatives, neighbours, or even strangers can’t poke through our medical records without our permission, we should have a right to expect that people can’t snoop through our genes,” says Kathy Hudson, who heads the Genetics and Public Policy Center in Washington DC.

Our experimental genome hack began like this: Peter drank water from a glass, which he handed to Michael. Michael’s first task was to get Peter’s DNA off the glass and turn it into a sample that he could submit to a genome-scanning company.

Michael approached several firms that ordinarily extract DNA from items

like drinking glasses and match this DNA against particular individuals, on behalf of the police, private detectives or citizens pursuing their own investigations. He said nothing about his intentions, but soon found a company that would extract the DNA without performing any DNA matches. Some weeks later a vial containing a solution of Peter’s DNA turned up at Michael’s home.

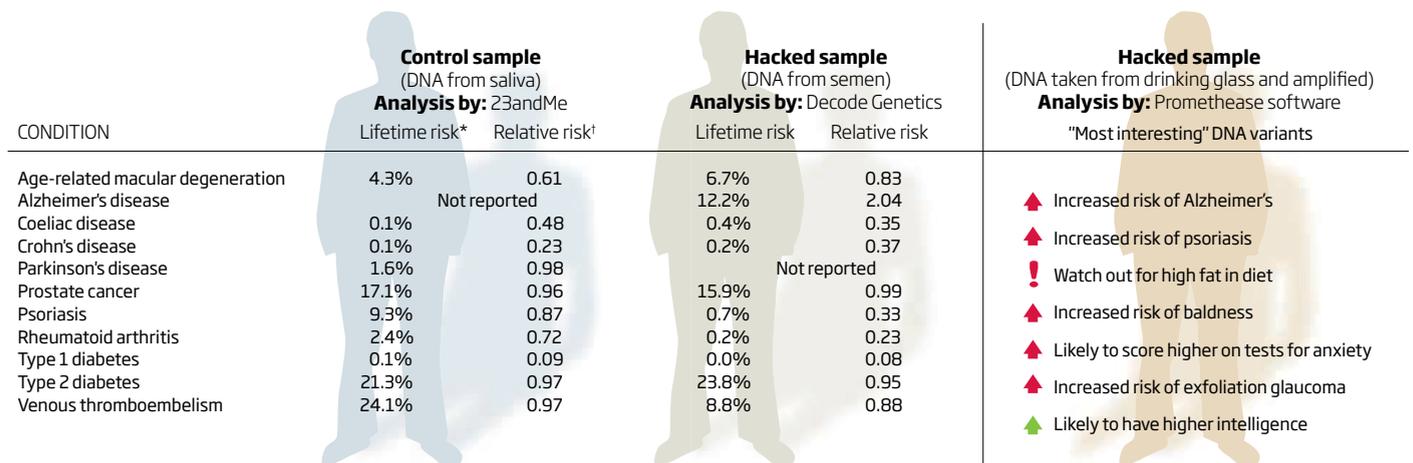
DNA boosters

Companies that perform genome scans use DNA “chips” that test for the presence of hundreds of thousands of DNA variants known as single nucleotide polymorphisms, or SNPs – some of which have been associated with susceptibility to various diseases. As these chips require more DNA than came from our drinking glass, Michael’s next challenge was to duplicate Peter’s DNA to get a large enough sample.

This procedure, called “whole genome amplification”, is offered to scientists and could, for instance, be used to amplify DNA from small clinical samples in studies investigating the genetic origins of disease. Geneticists often place orders involving large numbers of samples, but Michael found a lab services firm that was willing to amplify our

One genome, three views

If a genome hacker stole someone’s DNA, what might they learn about their victim? To investigate, we compared the interpretations of three genome scans of different samples, all taken from our reporter Peter Aldhous



* Peter’s chance of developing the condition in his lifetime [†] Peter’s risk, relative to the average for European ancestry (1.00 = average risk)

single sample to produce more than enough DNA to run on a SNP chip. He did not say why he wanted this done.

Next we had to choose a company to perform the genome scan itself. Lab services companies, such as the one that performed our amplification, often offer this service to scientists as well. But they do not provide an interpretation of the scans in terms of health risks and other traits – something a genome hacker is likely to want. So this wasn't our first port of call. Instead, we looked at the personal genomics services offered to members of the public by companies such as Decode Genetics of Reykjavik, Iceland, and the Californian firms 23andMe of Mountain View and Navigenics, based in Foster City.

Swab spiker

To gather the DNA provided by their customers, 23andMe and Navigenics use a collection tube into which you must spit about 2 millilitres of saliva. We decided that it would be hard to convert Peter's amplified DNA sample into a form that closely mimicked saliva. So we chose to use Decode's service, branded deCODEme, which instead collects DNA using swabs consisting of a piece of filter paper on a plastic handle that customers are supposed to rub against the inside of their cheek. We reasoned that Michael might be able to "spike" these swabs with Peter's amplified DNA without Decode noticing.

The terms and conditions for the deCODEme service state that someone submitting DNA must have the legal authority to do so, and that the sample must be taken from the cheek. We wanted to test whether deCODEme is vulnerable to abuse from someone prepared to ignore these terms, so Michael pipetted some of Peter's DNA onto deCODEme's swabs and sent them off for analysis under his own name. As far as Decode was concerned, it was a sample of Michael's DNA taken by swabbing his own cheek.

This is when we hit our only real obstacle. A few weeks later, Michael was told that the sample had not processed successfully. This is possibly because Decode uses a chip that isn't designed

to work with amplified DNA.

We had two contingency plans, however. First, Michael contacted our original lab services company again and asked it to analyse the remainder of our amplified sample using a different type of chip to the one that Decode uses. This company also has terms and conditions specifying that customers must have the necessary consents and approvals to submit samples. Mimicking a hacker who would be willing to ignore these terms, Michael submitted the amplified DNA for scanning.

Second, we made use of the replacement cheek swabs sent to deCODEme customers when a sample fails to process. We wanted to test the swabs' vulnerability to being spiked with a different source of "abandoned" DNA that might be taken by a genome hacker – semen from a used condom. Peter sealed the replacement swabs, spiked with his semen, in an envelope, which Michael sent back to Decode.

Both of these back-up plans worked. For the sample of DNA taken from the drinking glass and analysed by the lab services company we obtained a read-out of about a million of Peter's SNPs. To interpret this information, we used a computer program called Promethease, which can be downloaded for free from the genomics website SNPedia.com.

SNPedia contains information

Like the rest of us, our reporter Peter Aldhous leaves his DNA lying about – but it should still be private property

contributed by genomics enthusiasts on the diseases and traits linked to particular SNPs, mostly drawn from scientific papers. Promethease is a tool intended for legitimate customers of personal genomics companies that takes the raw data from an individual's genome scan and relates it to the information in SNPedia, highlighting those SNPs that seem to reveal the most interesting things about the person concerned.

For the semen sample submitted to Decode, we obtained the company's own interpretation of Peter's lifetime risks of developing a range of diseases, in addition to a full download of the raw data, again documenting about a million SNPs.

So what would a hacker who had taken Peter's DNA have learned about him? For the DNA taken from the drinking glass, Promethease highlighted a range of SNPs, including those conferring increased risks of baldness, the skin disease psoriasis, and a form of blindness called exfoliation glaucoma. Decode's interpretation of the semen sample was rather different. For instance, it decided from an analysis of eight different SNPs that Peter's risk of developing psoriasis is very low (see table, page 7). And while Promethease and Decode both concluded that Peter is more likely than a typical person to develop Alzheimer's disease, they disagreed on the size of his risk (see "A short-lived Alzheimer's scare", opposite).

In part, these confusing results reflect current limits to geneticists' knowledge of how individual variations in DNA sequences influence health. But the science is advancing quickly, so there is no room for complacency about the ease with which a genome can be hacked.

Motives for such hacking are not hard to find. In the wake of the US presidential election, Robert Green and George Annas of Boston University speculated that future campaigns could be blighted by the sneaky analysis of a candidate's DNA by political opponents who hope to reveal looming health problems (*The New England Journal of Medicine*, vol 359, p 2192).



MARK RICHARDS FOR NEW SCIENTIST

For people who are not politicians or celebrities, the most obvious threat comes from unscrupulous employers or insurers – and many countries have already restricted their use of genetic information. But private citizens may also have motives to pry into one another's DNA. A newly engaged person might want to know whether their future spouse carries genes making them vulnerable to dementia, for example. Or a childless couple could simply wipe a dribbling baby's mouth to investigate the child's genetic heritage and traits before deciding whether to adopt.

Cost is not a huge obstacle, as the sums we spent would not deter a wealthy snoop. Decode's analysis of Peter's semen cost \$985, while the total price for extracting his DNA from a drinking glass and then getting it amplified and analysed by the lab services company was about \$1700. Genomic analysis is only going to get cheaper, and more powerful. "The plummeting costs of genome profiling and sequencing make it all too tempting to snoop around in other people's genomes," says Hudson.

Still, the results of our investigation suggest steps that companies could take to help protect people's privacy, and *New Scientist* has informed firms that run SNP analyses of our findings.

For companies selling genetic analyses to the public, verifying the origin of samples will always be difficult unless sample collection is supervised by a medical professional or some other official witness. It is possible to run lab tests that distinguish saliva and swabs taken from inside the cheek from other biological samples, however.

Companies offering services to research scientists, meanwhile, might consider running some checks to try to confirm that customers are legitimate. Such checks may not be completely hacker-proof, but had Michael been asked for evidence of affiliation to a scientific institution, he would not have been able to provide it legitimately.

Following our investigation, the company that amplified and analysed the sample from the drinking glass is now considering whether it could

A SHORT-LIVED ALZHEIMER'S SCARE

We have shown that a genome hacker could take someone's DNA and obtain scans that reveal their risks of certain diseases (see main story). But how accurate are these scans, and how meaningful are the interpretations drawn from them?

To get an idea, we compared the scan results for three samples of DNA taken from our reporter Peter Aldhous. One scan was obtained legitimately by Peter submitting a sample of his saliva to the personal genomics firm 23andMe; the other two were from simulated genome "hacks". The first of these hacked samples was semen from a condom, submitted to a rival service provided by Decode Genetics; the second consisted of DNA extracted from a drinking glass, which was then amplified and scanned by a lab services company.

The raw data from

these scans, which document DNA variants known as SNPs, were reasonably consistent, according to an analysis performed for *New Scientist* by Kevin Jacobs, who runs Bioinformed Consulting Services in Gaithersburg, Maryland.

The raw data for the hacked semen sample were the same as for the legitimate saliva control for 99.996 per cent of the SNPs recorded in both cases. Meanwhile, the SNP data for the DNA taken from the drinking glass diverged a little from the results coming from the semen sample and the control, agreeing about 93 per cent of the time in each case. Why the glass sample gave slightly different results is unclear, but it might be due to degradation or contamination of the DNA, or artefacts introduced during its amplification.

Interpretations of the raw SNP data varied much

more widely, however (see table, page 7). Most confusing – and initially rather scary – were the suggestions about Peter's risk of developing Alzheimer's disease.

LOOK AGAIN

Alzheimer's risk is determined partly by variants of a gene called *APOE*. 23andMe provides no information on these variants, and for the other two samples we received conflicting interpretations.

For the DNA sample taken from the drinking glass, the software that we used to interpret the scan, called Promethease, highlighted a rare form of one SNP, which nestles close to *APOE* and tends to be inherited along with two copies of the risky variant of the gene, known as epsilon 4. Based on this, Promethease suggested Peter's risk of developing Alzheimer's disease was between 15 and 25 times that of

an average person.

It all seemed very worrying, until we looked at Decode's analysis. The raw data confirmed that Peter does carry this rare SNP, but Decode does not assess *APOE* in the indirect way that Promethease does. Instead, it analyses two SNPs in the *APOE* gene itself that define how many copies of the risky variant are actually present. This revealed that Peter has just one copy of epsilon 4, and one of the common variant, epsilon 3. On this basis, Decode concluded that he is only twice as likely to develop Alzheimer's as a typical person – a moderate risk that Peter shares with about 1 in 5 people of European descent.

So while the raw data from genome scans – legitimately obtained or not – are reasonably accurate, determining what they mean is another matter entirely.

introduce further checks without obstructing legitimate orders. "Clearly we do not want to process samples where the proper consent has not been obtained," says the firm's operations director. "It's a question of how to achieve that goal without impeding the research of legitimate scientists."

Thwarting genome hackers may also require new laws to protect privacy. One approach would be for other countries to follow the UK, which has made it a crime to have someone else's DNA with the intent of analysing it without consent. "Although we are not aware of any instances of this in personal genome analysis, there is a clear rationale for making it illegal to analyse an individual's DNA without their knowledge and consent," says

Decode spokesman Edward Farmer. Such laws are difficult to enforce, however, as an earlier *New Scientist* investigation revealed (31 January, p 6).

Another approach, which could be tried in parallel, would be to make it illegal for companies to extract and analyse DNA left on everyday items, except under specific circumstances. "There's no good reason, unless you are a police officer investigating a crime, to be doing DNA analysis on a sample from a drinking glass," argues Mark Rothstein, director of the Institute for Bioethics, Health Policy and Law at the University of Louisville in Kentucky.

One thing is clear: if lawmakers fail to rise to the challenge posed by genome hacking, we all have reason to fear for the security of our DNA. ■

"An engaged person might want to know if their future spouse is vulnerable to dementia"