9

The "discovery" of the "gay gene"

In 1993, the West was told that a scientist had discovered a "gay gene"—a gene causing homosexuality. The details were confusing for non-scientists, but the headline stuck. For Mr and Ms Average Citizen, it seemed that homosexuality might be genetic.

Actually there was no "gay gene." Even the scientist referred to, a gay man, Dean Hamer of the United States National Institutes of Health, never claimed to have found a gene determining homosexuality. "We have not found the gene—which we don't think exists— for sexual orientation," he said.[1] However, he claimed to have found evidence that some male homosexuality was passed through female members of a family. More specifically, he claimed to have found a linkage between homosexuality in males and a small stretch of the DNA on the X -chromosome.[2]

This chapter will look at these studies, but as discussed in Chapters One and Eight, scientists now believe that thousands of genes may be involved in almost any trait and that gene expression depends on environmental events and even social interactions. Gene patterns may be a recipe for tissues and bodies, but don't dictate behaviours. Though much research has tried to find specific SSA genes, none have yet been conclusively found. Any connections are very weak, indirect, not specifically

sexual and we'll see that a very large 2019 study shows an alarming amount of early work was simply wrong.

Gene linkage studies

Hamer's work falls into a category of research called "gene linkage studies." There was a surge of research in this field in the late twentieth century but because thorough "whole genome" scans are now the norm, gene linkage studies are becoming rather passé. A whole genome scan means all the genes are examined; a gene linkage can only look at a few at a time.

The first most spectacular linkage study, was the discovery, early in 1993, of a gene responsible for Huntington's disease. The gene had already been tracked down to chromosome 4, but it took six teams of workers at ten different institutions ten years to find whereabouts on chromosome 4. Over the succeeding decade, researchers also identified genes causing cystic fibrosis, muscular dystrophy, and other diseases.

From 1990 to 1993 biologists had astonishing success mapping the human genome (on schedule and within budget!) and analyses are still being published. In one five year period near the end of the nineties, the genes corresponding to 1450 *physical* conditions were identified and their precise location on various chromosomes found. Inspired by these successes, some scientists began talking optimistically of uncovering the genetic basis to human behaviours in the same way. This is what Hamer tried to do, and what other scientists, called behavioural geneticists, had attempted to do before him, but with scant success.

What happens in Gene Linkage studies?

In linkage studies for behaviour, researchers look for an extended family with an unusually high incidence of some behaviour, such as bipolar disorder, and then take samples of tissue from all available members and analyse the DNA, looking for segments in common using sets of tiny, synthesised DNA segments, called "markers"—an identical set for each person. These tiny markers are configured in such a way that they attach in a lock and key fashion to any stretches of DNA that mirror the markers; they usually contain a small range of genes. Searching for one gene in 22,000 is worse than looking for a contact lens in a swimming pool, but, in this way, segments of DNA (also containing "irrelevant"

genes) can be found in different people. If the same sequence is associated consistently with a given trait, then researchers assume the marker lies close to the gene that codes for it, along with the other irrelevant genes. At that point, the researchers believe they have found a linkage.

The strength of linkage analysis is in studying *physical diseases* that have distinct symptoms and are caused by a single dominant gene. When they attempt to link *behaviours* to a single gene, they run into a volley of scientific scepticism, for several reasons.

First, no mainstream geneticist believes that behaviour is linked to one single gene (see Chapter One). "It's very rare to find genes that have a specific effect," says Harvard biologist Balaban.[3] Second, in the word of one writer for *Science*, "the field of behavioural genetics is littered with apparent [gene linkage] discoveries that were later called into question or retracted."[4] It was only in the first decade of the 21st century that gene linkage studies became more reliable. Unfortunately the supposed SSA—genetic link was publicised before that time. And, as mentioned, the most recent studies have moved beyond linkage studies to very detailed scans of the entire genome.

In the next section we survey gene linkage studies that have tried to identify genes linked to schizophrenia, to put in perspective what is needed for success in gene linkage studies.

About the time Hamer sought to associate SSA with a section of the X-chromosome, linkage studies were scientifically dubious, but seemed worth pursuing although similar gene linkage studies on schizophrenia and alcoholism had given rather contradictory results.

Schizophrenia

Gene linkage studies on schizophrenia blossomed with the completion of the human genome project. Using markers, many regions were found on various chromosomes which correlated strongly with schizophrenia, and studies on fresh family lineages and families from other ethnicities often confirmed them, though there were puzzling lacks of confirmation from time to time.

However the results for some regions of the DNA seemed so convincing that scientists began looking for specific genes within them. By August 2005, at least 25 chromosome regions were thought to be involved, and an equal number of genes on them were being investigated.

Of these there was strong evidence for involvement of 4 genes and "promising but not compelling evidence" for a fifth. Some of the results were described as "very robust." This was a good consensus to emerge from a welter of initially inconsistent gene linkage studies. The work had progressed so far that some researchers started to experiment with drugs which interacted with the products of the genes known to be involved, in the hope of reversing or at least reducing the progress of schizophrenia.

But this confidence proved to be completely ill-founded. By mid-2010 "whole genome" scanning had thrown the gene linkage results into embarrassing disarray. In "whole genome" scanning—rather than using markers which result in rough screening only—all the genes are scanned in extraordinary detail, nucleotide by nucleotide. Nucleotides are subunits of DNA There are hundreds of nucleotides in a single gene, each made up of a nitrogen base, a sugar and phosphate.

Enormous multicenter efforts scanned the entire genomes of 7662 subjects and 29053 controls in one study alone; a second involved 3322 subjects and 3587 controls, and a third involved 8008 subjects and 19077 controls but altogether they could not confirm *any* of the previous gene-linkage work, only labelling them promising. The detailed saga is recounted elsewhere.[5] This was embarrassing because so much previous work now seemed premature. One million gene variants were examined, involving most common variations of DNA nucleotides. They found absolutely unequivocal evidence of a connection to variants in a gene on chromosome 6 linked to immunity, and to three other completely new genes, two called transcription factors (TCF4 and ZNF804A, the latter a "zinc finger" protein because of its composition and shape) and the last, called neurogrannin, but, disconcertingly, *noone* had previously suspected them of being involved. The transcription factors were used by the nucleus to read the DNA sequence and neurogranin is a brain-specific protein connected with biochemical control of calcium. Like the fruit-fly case we described in Chapter One, why these genes should be important in schizophrenia is not at all obvious, and links will be very indirect.

Schizophrenia is certainly reliant on multiple genes, because four genes were found and others suspected: but these significant genes found only account for 3% of schizophrenia. The effect is weak. This is a vivid illustration of how difficult this field is.

Hamer's Study—SSA

Compared with the scale and outcomes of the schizophrenia project above, early efforts which attempted to link genes with SSA now seem embarrassingly small, very naive and hyper-optimistic. Moreover, Chapter Ten shows the genetic contribution to SSA calculated another way is relatively low, lowering the prospects of success from gene studies.

However: To find the homosexual gene or genes, Hamer and his colleagues[2] first recruited 76 men, who identified themselves as predominantly or exclusively homosexual. They found 13.5% of their brothers to be gay, much higher than the 1% occurrence of exclusive homosexuality in the general male population, and also found a higher level of homosexuality in maternal uncles and the sons of maternal aunts. They then recruited 38 families in which there were two homosexual brothers, suspecting this would show more clearly the effect of homosexuality and Hamer searched for a linkage on the X (female) chromosome.[2]

Hamer claimed to have found a "statistically significant correlation" between the homosexual orientation and a genetic sequence on the tip of the long arm of the X chromosome, an area called "Xq28". Hamer published his paper in *Science*, in July 1993, and immediately became a controversial figure in the scientific community. Numerous letters to the journal *Nature* were mostly critical.

In the meantime, Hamer[11] and colleagues replicated their study using a new population. This time, the results were less impressive—only just statistically significant, but the replication was promising and reassuring.

Hamer's study on the "gay gene" was then contradicted in a gene linkage study[12] published in Western Ontario, headed by researcher Rice. Rice found no trace of an association between homosexuality and the genetic region Hamer and his team had pin-pointed. Even when the results from all the Hamer and Rice studies were combined, there was no significant association. Hamer argued that the Rice team result was inadequate because they did not select homosexual men with an excess of maternal homosexuality.

Then a "whole genome" study[13] appeared from the National Institutes of Health in Maryland, with collaborators from several parts of the US. It was much larger than any preceding gene linkage study.

The first author was called Mustanski, and Hamer was included in the author list, though not leading the study.

According to the results in the paper, *no* part of the entire genome was statistically significantly linked with SSA. One peak on Chromosome 7 (region 7q36) approached statistical significance but the result did not survive replication by a 2014 study.

Then, using a different method, the Rice team[10] could not replicate the Mustanski results. So, more conflict!

In mid 2014 a Chicago researcher called Sanders headed a team which published[8] the result of investigating the genetic links yet again, working on a sample of 409 SSA brothers. They found more convincing confirmation of the Xq28 linkage, but only suggested specific genes which might be involved. Their comment is worth citing, "We also emphasize that genetic contributions are far from determinant but instead represent a part of the trait's multifactorial causation both genetic and environmental." Translation: genes as a whole are a minor contribution; there are many factors involved.

Much earlier Hamer's group attempted an SSA-gene linkage study on lesbians but did not find a link between parts of the X-chromosome and the presence of lesbianism in families.

A 2015 Chinese study showed a connection between a gene called COMT and sexual orientation,[7] but calculation shows the effect size is weak.

The large 2019 genome/SSA study

In 2019 the results of a very large study appeared in *Science,*[14] one of the top scientific journals, which claimed discovery of five genes connected to SSA. They paid careful attention to statistical validity and the gene discovery is probably correct, but their definition of SSA is surprisingly poor, and the connection doesn't mean very much. Perhaps they will publish better material in future.

Where can you get the tens of thousands people needed for such a gene/SSA study? Today it comes from places many readers will have patronised—the half dozen companies analysing DNA for private clients. Most results came from the UK Biobank company; nearly 409,000 volunteers had agreed to a survey on sexual matters. Results added in from

other related companies and surveys increased this to 477,000—nearly half a million. We'll call this the Biobank study.

More than twenty authors are listed: from the USA, Sweden, Denmark, Netherlands, UK, Australia and a combined research group from the USA 23-and-Me genome company. This is Big Science.

The most serious problem is that researchers divided the group into two classes: those who never had a same-sex partner, and those who had at least one. Previous surveys describe this as a mediocre classification.

Even Kinsey in the late 1940's talked about those who had merely incidental SSA experience: one or two experiences and nothing thereafter. That's the present case. Researchers know very well that many of these sorts of encounters are exploratory or even sexual abuse, and not a continuing sexual orientation. In fact, in this study, they comprise most of those with some same-sex attraction. Laumann et al. (Chapter Two) found 7% of men had reported one or more same-sex partners but those active at the time of his survey were only 2.9% and exclusive SSA men were about 1%. Most had not persevered. In the same way Laumann et al. found 4% of women had one or more same-sex partners, but those active at survey time were only 1.8%. This means the Biobank study is *mostly about sexual explorers.* It's dubious practice to label them all "homosexual".

The researchers warned there were two qualitatively different classes of people—those slightly non-heterosexual, and those exclusively homosexual. The volunteers overall have a rather weak same-sex drive. Why didn't researchers concentrate on those with a strong drive? Well, that would probably have reduced their sample size by a factor of 10, which would make the results much less clear. So they faced a trade-off between mediocre sample description or mediocre statistical test power.

"Getting your DNA done" is quite popular and perhaps the survey will be repeated when there are ten times as many people available and a large active SSA group, but let's see what was possible even with the available sample and doubtful sexual classification.

The researchers present the Biobank results first, and for men they found a connection between four genes and some SSA experience. Embarrassingly, these genes had never been implicated in nearly a dozen similar preceding studies probably involving several million dollars of effort. All the previous work was useless because samples were too small,

but this was realised clearly only in the last five years or so. Even more embarrassingly, the controversy about the genes on the X-chromosome, particularly the XQ28 region was pointless—none of the four genes Biobank researchers found were on the X-chromosome.

For the very first time researchers found three genes correlated with SSA in women, and two of these were also found in men. No previous work had found any gene connections for women. There was some overlap then, between genes for men and women and SSA, but overlap between men and women for most unrelated traits in other studies was much higher. Could SSA be partly different in men and women? Quite reasonable.

When the researchers checked the results using much smaller samples from other sources, and a total of 15000 individuals, they confirmed three of the results, which is a good test of reliability, but the Biobank large sample results were far more reliable.

Two of the genes were connected to smell sensors. Could this be SSA related? But previous studies could also point to vague connections between their spurious genes and various functions and were wrong. So even present alleged connections should be treated rather sceptically.

At this point you may be thinking, "Well, there may not be one unique gene, but a handful. OK, so a small cluster of genes are responsible for SSA? And they have a powerful effect?"

No, they don't! The researchers were able to calculate the strength of any effect, and an individual with one of the four genes is at most 0.4% more likely to be SSA. Yes, almost negligible. But it is typical of what gene researchers find, which is why they conclude that many, many genes influence traits, each with a *very* small effect strength. For the Biobank study, the researchers were able to show that the minute influences were spread fairly evenly among all the chromosomes, again confirming there were very many genes and on all the chromosomes.

But what was the sum of all these many small influences? The researchers were able to calculate a range depending on various assumptions and it was 8-25%. In the paper they imply a typical estimate of the total influence strength would be 10%—as derived elsewhere in this book. If 0% is no influence, and 100% is a dictatorship, then 50% would be a medium influence, but 10% is quite weak—and obviously quite indirect.

If there was a strong physical effect on SSA, you'd expect special genes concentrated in parts of the body, maybe within the brain or in the sex organs. The researchers tested individual tissues for other genes suspected of some correlation with SSA but didn't find it, in fact they found very few correlations with other physical traits (an exception was a finger length ratio in women).

The researchers identified openness to new experiences and risk behaviour in their group of people who had at least one same sex encounter in the last year—though this was not a genetic test. But it again raises the issue of whether their study was of people with SSA or of sexual explorers.

That could also account for the partial gene similarities (overlap) between the men and women—i.e. the common factor is openness to new experience.

So, the Biobank Study, though impressive in its reach and resources, is limited in reaching conclusions about genetic effects on SSA.

Summary

The authors of the paper also strongly emphasise a DNA test for gayness is not possible. The scientific community realises that "our genes do not make us do it". Hamer has always believed that. To give him the last word: "There will never be a test that will say for certain whether a child will be gay. We know that for certain."[9] This means as clearly as anyone could state, that no-one is born gay.

Those who believe that homosexuality has psychological and sociological explanations have no difficulty with the possibility of genetic linkages to homosexuality. They would argue that any genetic link to a physical characteristic that might heighten a person's sense of gender non-conformity (a strong known predictor of later homosexuality), could be held to be a contributing factor to later homosexuality. In a boy these might be, e.g. genes related to slightness of build or poor physical co-ordination (making a boy poor at sports). In a girl they might be factors like atypical physical strength, shape, height, or weight, or a more masculine finger-length ratio. Links? Yes, but weak and indirect.

References

1. McKie R. 1993. The myth of the gay gene. *The Press*(30 July):9
2. Hamer DH, Hu S, Magnuson VL, Hu N, Pattatucci AML. 1993. A linkage between DNA markers on the X-chromosome and male sexual orientation. *Science* 261:321-7
3. Horgan J. 1993. Eugenics revisited. *Scientific American* 268 (June):92-100
4. Pool R. 1993. Evidence for homosexuality gene. *Science* 261:291-2
5. Nieratschker V, Nothen MM, Rietschel M. 2010. New genetic findings in schizophrenia: Is there still room for the dopamine hypothesis of schizophrenia? *Frontiers in Behavioral Neuroscience* 4:23-32
6. Plomin R, Hill L, Craig IW, McGuffin P, Purcell S, Sham P, Lubinski D, Thompson LA, Fisher PJ, Turic D, Owen MJ. 2001. A genome-wide scan of 1842 DNA markers for allelic associations with general cognitive ability: a five-stage design using DNA pooling and extreme selected groups. *Behavior Genetics* 31:489-95
7. Yu, W, Tu, D, Hong, F, Wang, J, Liu, X, Cai, Y, Shu, R, Zhao, G, Wang, F, Pan, H, Wu, S. (2015) Analysis of the Association between Catechol-O-Methyltransferase Val158Met and Male Sexual Orientation. *Journal of Sexual Medicine,* 12(9):1920-1926
8. Sanders, AR, Martin, ER, Beecham, GW, Guo, S, Dawood, K, Rieger, G, Badner, JA, Gershon, ES, Krishnappa, RS, Kolundzija, AB, Duan, J, Gejman, PV, Bailey, M 2014. Genome-wide scan demonstrates significant linkage for male sexual orientation. *Psychological Medicine* 17:1-10
9. Holmes B. 1994. Gay gene test 'inaccurate and immoral'. *New Scientist* 141 (5 March):9
10. Ramagopalan SV, Dyment DA, Handunnetthi L, Rice GP, Ebers GC. 2010. A genome-wide scan of male sexual orientation. *Journal of Human Genetics* 55:131-132
11. Hu S, Pattatucci AML, Patterson C, Li L, Fulker DW, Cherny SS, Kruglyak L, Hamer DL. 1995. Linkage between sexual orientation and chromosome Xq28 in male but not in females. *Nature Genetics* 11:248-256
12. Rice G, Anderson C, Risch N, Eber G. 1999. Male homosexuality: absence of linkage to microsatellite markers at Xq28. *Science* 284:665-7
13. Mustanski BS, DuPree MG, Nievergelt CM, Bocklandt S, Schork NJ, Hamer DH. 2005. A genome-wide scan of male sexual orientation. *Human Genetics* 116:272-8
14. Ganna A, Verweij K, Nivard M, Maier R, Wedow R, Busch A, Abdellaoui A, Guo S, Sathirapongsasuti J, Team 23andMe Research, et al. 2019. Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior. *Science* (80-). 365:eaat7693.