Schizophrenia is a devastating and chronic neuropsychiatric disorder that affects nearly 1% of the world’s population. It has long been hoped that the identification of genetic risk factors for schizophrenia would help to identify the physiological causes of the disease, but, despite decades of intensive research, the biological underpinnings of schizophrenia have remained elusive. On page 177 of the issue, Sekar et al. present a remarkable genomic and neurobiological study that finally delivers on this long-standing hope.

There are two typical paths to drug discovery. The first is serendipity — sometimes researchers simply get lucky and find compounds that happen to work. Although such fortunate accidents have led to the development of some of medicine’s most important drugs, this method of discovery cannot be systematized, and the resulting treatments often fall well short of a cure. Current treatments for schizophrenia fall uniformly into this category. The second, preferable, path is to understand the mechanisms that underpin a disease, and to design treatments that target those underlying causes.

One of the main problems in deciphering the molecular basis of schizophrenia is that there is a near complete absence of clearly associated biological changes. However, schizophrenia is a strongly genetic disorder, and for decades many have believed that understanding the genetics involved might provide a way to begin to dissect the biology of the disease. Until now, following this line of reasoning has been a largely frustrating experience. Before the advent of genome-wide association studies (GWAS), which link disease risk to specific regions of the genome, the field was plagued by a lack of statistical rigour and false starts, as researchers reported hundreds of putative associations between genetic variations and the risk of schizophrenia. But GWAS have demonstrated that these early associations were largely incorrect.

Careful GWAS have now finally identified real associations between genetic variants and schizophrenia. In particular, a 2014 study that used a huge sample size and strict statistical methods revealed that there are at least 108 genetic regions (loci) associated with a risk of schizophrenia. However, finding a locus is not the same as identifying a causal gene — for instance, loci are often located in sequences that don’t encode genes, and many different variations can occur in one region, making it hard to pinpoint what exactly is driving the risk signal. Indeed, these 108 risk factors were not traced to specific genes or variant sequences. Without that information, a clear insight into the molecular aetiology of the disease remained lacking, but Sekar et al. have now taken that crucial next step.

The strongest risk association for schizophrenia is found in the major histocompatibility complex (MHC) — a region on chromosome 6 that contains genes involved in acquired immunity (Fig. 1). Within the MHC, the strongest risks are associated with loci near the gene C4, which encodes a complement factor (a part of the innate immune system). This association motivated the authors to ask whether variations in C4 might be involved in schizophrenia risk.

In humans, C4 exists as two distinct genes, C4A and C4B, and the number of copies of each gene varies from person to person. To complicate things further, there are long and short versions of C4A and C4B, which are determined by the presence or absence of a human endogenous retroviral (HERV) insertion in a non-coding region of the gene. Sekar et al. developed a way to accurately assess the number of copies of each gene, and then used RNA-expression data to relate gene copy number and HERV status to C4A and C4B gene-expression levels.

With these data in hand, the authors present compelling evidence that they have identified the underlying cause of one of the main MHC risk associations. They have found that variation at C4 is associated with schizophrenia risk independently of other genetic variations in the MHC region. Specifically, different levels of risk are associated with different genetic combinations of C4 copy number and HERV status. The authors used RNA data to show that expression of C4A and C4B increases with copy number, and that the presence of the HERV insertion increases the ratio of C4A to C4B expression. Moreover, the higher the levels of C4A expression, the greater the risk of schizophrenia.

It is worth emphasizing that defining causal variants in the MHC regions is one of the most challenging problems in human statistical genetics. The MHC is 3.6 million bases long, and variants associated with schizophrenia risk have been found across the entire region. Indeed, the region is so challenging that geneticists often joke that they hope to find associations anywhere except in the MHC. Caution is imperative when trying to pinpoint causation in such a genetically complex region. However, Sekar and colleagues seem to have overcome the challenges of studying the MHC to pinpoint how variations in the C4 genes that alter gene expression can increase risk.

**Figure 1 | A complex association.** Variations in numerous regions of the genome have been associated with an increased risk of schizophrenia, with many of the strongest associations occurring in the major histocompatibility complex (MHC). This region is challenging to study because it is 3.6 million bases long and complex, containing three classes of gene that have differing roles in immunity. Genetic loci that are associated with risk are found along the entire region (only three of the strongest associations are shown here, for simplicity). However, genetic variants in C4A and C4B are most strongly associated with increased risk. Sekar et al. overcome the challenges of studying the MHC to pinpoint how variations in the C4 genes that alter gene expression can increase risk.
The complement pathway has been implicated in synaptic pruning, a developmental process in which the synaptic connections between neurons are continuously eliminated in the brain until early adulthood. Using a mouse model, Sekar et al. found that C4 expression is upregulated during periods of synaptic pruning. By contrast, mice deficient in C4 showed signs of decreased pruning. Thus, the authors postulate that increased C4A expression in individuals with schizophrenia results in increased synaptic pruning. Interestingly, studies of the brains of humans with schizophrenia have shown that affected individuals exhibit thinning and reduced synaptic structures in the cortical region of the brain compared with people without the disorder. Hyperactive synaptic pruning might explain these findings.

Unfortunately, because mice lack the two forms of C4 found in humans, the question of why schizophrenia risk depends specifically on C4A expression levels remains open. Answering this question is now a priority, and can be expected to provide further mechanistic information.

Sekar and colleagues’ study finally gives us the first real inroad into the molecular aetiology of schizophrenia, and perhaps a direction for the development of treatments. Although pruning undoubtedly represents a challenging therapeutic target, the authors’ beautiful and comprehensive study gives much-needed inspiration for all those researchers who are trying to leverage genetics to advance our understanding of the biology of neuropsychiatric diseases.

This key set of observations can be explained only by the presence of a single, widespread ice shelf (Fig. 1), one that flowed freely between large ridges, but became grounded and left its trail across the shallower regions of the Arctic basin. The ice shelf — estimated at some 4.1 million square kilometres on the basis of the numbers given in Jakobsson and colleagues’ paper — was four times the area of all the ice sheets now found in Antarctica, including the immense Ross and Ronne–Filchner ice shelves. It must have been about 1 kilometre thick, allowing it to float over the great depths of the Amerasian and Eurasian basins, but touching down and ploughing across the shallower regions.

By tracing the glacial furrows towards their points of origin, Jakobsson and co-workers conclude that land-based ice sheets must have fed the ice shelf from three separate centres: the eastern Siberian lowlands; the Scandinavian regions of the Barents Sea; and high Arctic islands in north central Canada (Fig. 1). Furthermore, the authors found evidence — in the form of microfossils and of magnetic properties in sedimentary cores collected from some