UK CFS/ME Research Collaborative conference report

13-14 October 2015

This report will be published in parts as and when they are ready.

This part covers the following presentations:

The bigger picture: genomics, epigenomics and other omics
Anne Faulkner Memorial Lecture by Prof George Davey Smith, University of Bristol

The bigger picture: genomics, epigenomics and other omics
Anne Faulkner Memorial Lecture by Prof George Davey Smith, University of Bristol

Prof Davey Smith presented some new and potentially powerful approaches to researching illnesses with unknown causes. The technique uses genetics and genomic data to discover non-genetic, modifiable causes of disease and disease progression.

Genomics looks at common variations in genes, across our whole genome, and attempts to relate these to traits and conditions. These studies have been possible due to the sequencing of the human genome.

Prof Davey Smith began the presentation with some existing examples of CFS/ME genetic studies, to illustrate the need for much larger sample sizes. He picked out several papers at random with small samples (12-76 patients) that were too small and likely to be false positives.

The statistical power of these very small studies is limited because, in order to rule out the possibility of false positives and the results happening by chance, much larger sample sizes are needed.

Genomics

Genomics studies look at common germline genetic associations across the genome, relating them to traits and conditions. This method of research has really expanded in the past few years due to the sequencing of the human genome.

Up until 2005, this field of genetics had similar problems as current CFS/ME research, in that studies with very small sample sizes were carried out, producing false positives and an inability to replicate results. Prof Davey Smith co-authored a paper which suggested that virtually all of the genetic association studies that had been published in journals up until 2002 had resulted in false positives (Colhoun et al, 2003).

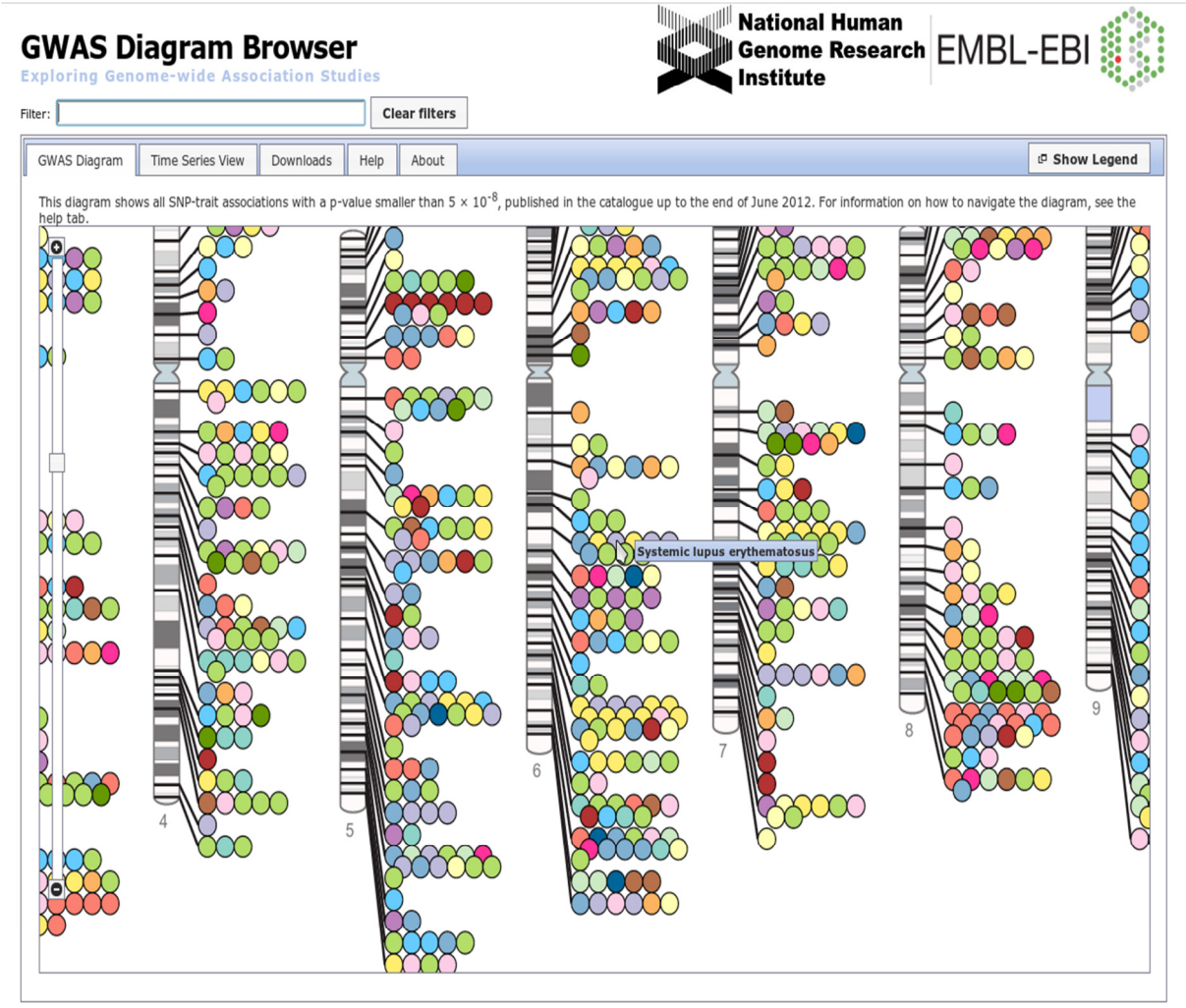
Inadequate sample sizes suggest that the most important factors underlying inability to replicate these associations are publication bias and failure to attribute results to chance.

These are problems that are all rectifiable. If we don't fix them, we risk wasting scientific effort and the rejection of a potentially useful research strategy. Thankfully, the situation has been transformed by groups getting together and carrying out much larger, more useful studies, driven by funders such as the Wellcome Trust.

It had been discovered that particular genetic variants are reliably associated with a large number of traits, including diabetes, obesity and cardiovascular disease. There are now more than 10,000 established, replicable findings.

Studies that examine if common genetic variants are associated with a disease are called Genome-Wide Association Studies (GWAS). The following diagram shows a small section of the chromosome from the GWAS Diagram Browser. The studies can be searched for a specific condition and each dot on the diagram represents a

section associated with traits, with systemic lupus erythematosus highlighted. There are currently no established reliable genetic associations with CFS/ME.



What do robust and reliable genetic associations tell us about the following?

1. Classification of disease

The difficulty with the classification of disease is that diseases often cannot be defined and classified clearly, especially when the aetiology or pathogenesis is unknown. Thus diagnostic terms often only reflect a set of symptoms, as with CFS/ME. Genetic associations can validate an established disease classification and predetermined subcategories of a condition.

2. Causal factors for disease

Germline genetic variation can be used to provide robust evidence of the effects of putative modifiable causal factors for disease without conducting a traditional randomised trial. This approach is known as Mendelian randomisation (MR).

Conventional observational epidemiology studies have provided unreliable evidence regarding causal factors. For example, a study of participants taking a vitamin E supplement suggested vitamin E reduced risk of coronary heart disease by 40%

(Rimm et al, 1993). Although studies such as this adjust for confounding risk factors, such as diet, weight, smoking and alcohol consumption, when later RCTs were carried out, they found no reduced risk of CHD from vitamin E compared with placebo.

The reason for this might be that taking vitamin E is strongly associated with other characteristics which influence CHD risk. This was confirmed by a study (Lawlor et al, 2004) which showed that people taking a vitamin E supplement were less likely to come from a poor background, less likely to smoke and be obese, and more likely to exercise, drink alcohol, and have a low fat diet.

MR can be used to get around confounding factors by using the genetic variation which relates to the factor of interest, in this design the genetic variants maybe referred to as “instrumental variable”. The MR method has been applied in a number of fields, which has both proved and disproved existing theories from observational studies.

To use vitamin D as an example: people naturally have different levels of vitamin D, and some of this variation is down to the versions of genes they have: some have genes that are better at making vitamin D. So you can select people by their vitamin D genes and group these into those predicted to have higher or lower levels of vitamin D (actually testing their vitamin D - related levels shows this technique works very well). The vitamin D genes are randomly spread through the population (this is MR), and not associated with confounding factors such as diet or social class. So this is a better way of testing the true effect of vitamin D.

Genotypes therefore can proxy for some modifiable risk factors, and there should be no confounding of genotype by behavioural, socioeconomic or physiological factors, except those influenced by alleles at closely proximate loci or due to population stratification.

The 10,000 genetic traits in the GWAS catalogue include genetic variants related to health behaviours such as smoking (eg. nicotine receptor variants relate to the degree of which people become addicted to nicotine) and drinking alcohol (eg. some people are naturally intolerant of alcohol), along with genetic variants related to biomarkers.

3. Establishing cause and effect

Bi-directional MR can sort cause from effect. For example, there is an established connection between vitamin D and Body Mass Index (BMI) and so it was thought that vitamin D could lower BMI. MR can be performed bi-directionally, and it has shown that there is no effect of vitamin D on BMI but looking at the reverse, ie. the effect of BMI on vitamin D levels, MR shows that higher BMI leads to lower circulating vitamin D.

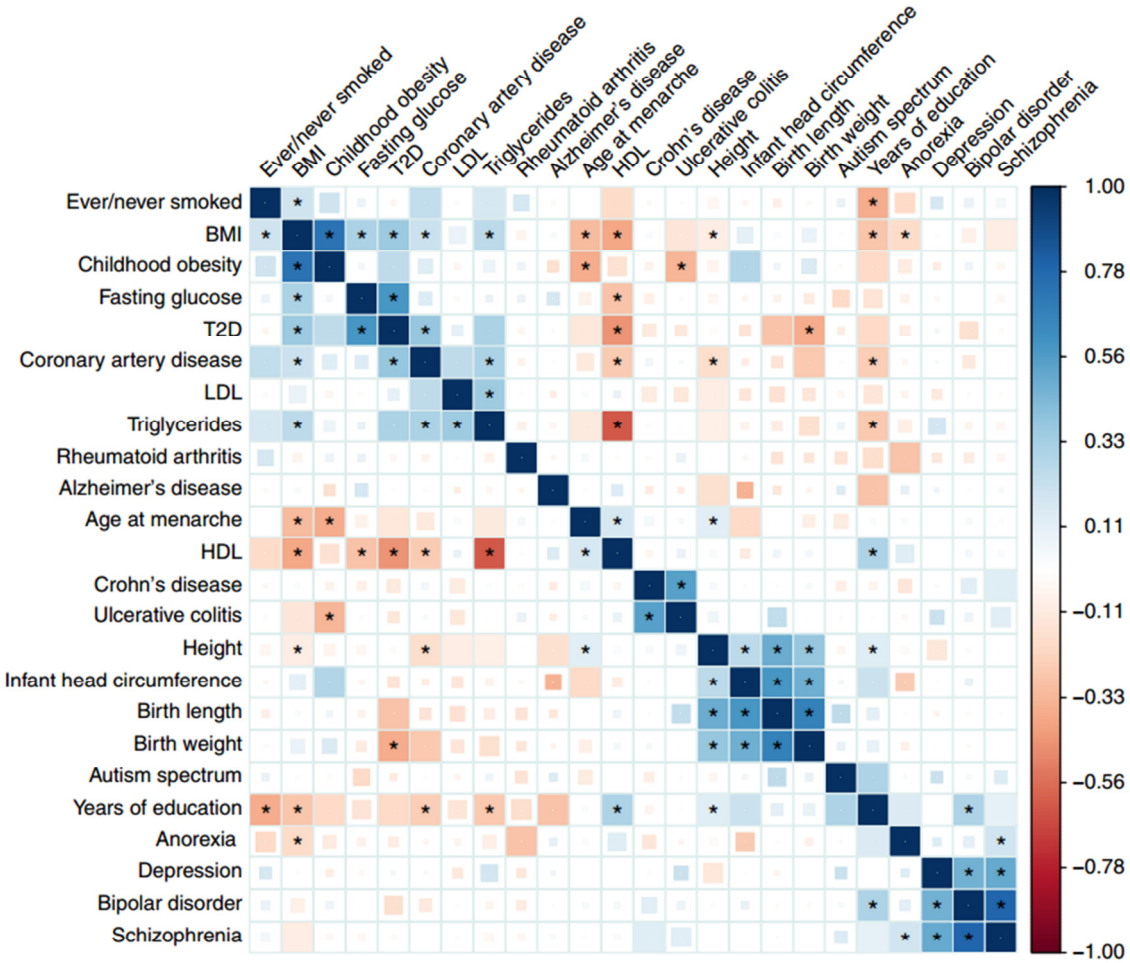
The advantage of MR studies are that a huge amount of genome-wide data is already available and can be analysed without performing expensive large RCTs.

4. Suggesting causes where aetiology is unknown

"Reverse MR" can suggest causes of specific diseases that have not been thought of. For example, large GWAS studies found an association of the genetic variant related to nicotine addiction with lung cancer. If it hadn't already been known that smoking causes lung cancer, this research would have pointed to smoking being a cause.

5. Genetic overlap of traits

Statistical techniques can show how traits are related, eg. Alzheimer's with low childhood BMI, which may suggest cause in one direction or another. If one trait caused another, a genetic correlation would be seen. Unlike MR this does not give the direction of the causal relationship, however.



Bulik-Sullivan B et al. An atlas of genetic correlations across human diseases and traits. Nature Genetics 2015

What does use of genetic association data not tell us about?

If the sample sizes are inadequate, these studies give unreliable results. This can be worse than having no data at all because such studies can mislead and result in more research being carried out based upon them, wasting research funds. Very large sample sizes, of the order of many thousands, are needed in order to produce reliable evidence.

GWAS do not necessarily tell us about treatment or cures. For example, if MR shows that smoking causes lung cancer, it does not show that stopping smoking treats lung cancer once it has developed. This is the case with many diseases, where once the disease process has been triggered, intervening in what triggered the disease no longer has any effect.

For any condition two studies are needed: a GWAS of having the disease versus not having the disease, to indicate what might cause it. Having a really well-established cohort of people with the disease to look at genetic variants related to disease progression is also required, to reveal information about potential treatments.

Other "omics"

1. Epigenetics

From a blood sample, we can look at epigenetic markers on the germline DNA which are sometimes related to whether that DNA is expressed or not (in over-simplified terms whether the gene is switched on or off). This might tell us how the environment influences or works together with the genome to develop disease or mediate the effect of the disease.

Epigenetic markers aren't protected from change, unlike germline/DNA, so they're affected by the same confounding factors as influence exposures in observational studies (like taking vitamins, as discussed above). DNA methylation is the most stable epigenetic measure, and it is the one which has been most studied.

Epigenetics is currently popular in the media, with examples such as epigenetic indicators of sexual orientation, and whether trauma is passed on in the genes of Holocaust survivors, having received considerable recent media attention. Small sample sizes were used in these studies and the findings are often wholly unreliable.

What we can learn from epigenetics is that processes can occur which lead to the same genome being expressed differently. The cells in our body all have the same germline genome. The cells start out the same but then during development they differentiate into different types of cells, such as liver cells and brain cells. Cell division leads to cells becoming specified where liver cells divide they produce more liver cells. Epigenetic processes are therefore key to development. In disease, this means that, as diseased cells divide, the disease is passed on to new cells.

Epigenomics provides ways of studying the mechanisms of gene-environment interaction and provides robust evidence of the experience and timing of some exposures, which could help understand mechanisms of disease. For example,

smoking during pregnancy affects foetal DNA which then remains evident until adulthood at least.

2. Other "Omics"

Gene expression is transient and the study of this (transcriptomics) has somewhat been disappointing. It was thought there would lead to reliable predictors of disease progression but so far there are few replicated findings.

Proteomics is the study of a wide range of proteins. This approach has considerable potential, but currently the technology is expensive and difficult to implement in very large-scale studies.

Metabolomics is the study of metabolites, and has considerable potential, as many metabolites (and proteins) are amenable to pharmacotherapeutic manipulation, so can be the target of treatments. Mendelian randomisation can be utilised using genetic variants that are reliably related to proteins and metabolites to investigate how these may be involved in the progression of disease, and therefore are appropriate targets for prevention or treatment. A further technology could be referred to as "immunomics", the study of a large number of molecules related to immune processes.

Recently platforms that measure large numbers of factors within the immune system are being developed and could be applied to large-scale studies, although currently they are expensive. Microbiomics is the study of the microbiome, as the name implies. This is a highly fashionable area and the headline figure that 9 out of 10 cells in the human body aren't actually human – they are the cells of microbiological organisms that inhabit the human body – is a popular statistic used to suggest that these organisms could have an important influence on human health. Currently there is more hype than solid evidence in this field, but clearly it remains of potentially high importance.

Prof Davey Smith ended his presentation by asking if CFS/ME research should be "business as usual" – i.e. many different groups carrying out under-powered small studies, producing often non-replicable findings - or if there was a better way ahead? He suggested a co-ordinated and collaborative approach and asked conference delegates to work together to find the best way forward for CFS/ME research.