

GWAS experimental design:

Based on “stage”, there are mainly two ways to design experiments. The first one is One-stage design and another is Two-stage design or Multiple-stage design. In One-stage design, we only select a large enough samples once, and genotype all of them. Then, we analysis the associated effects between each SNP and traits. In two-stage design or multiple-stage design, we usually select a small sample group for genotyping. Then, we select the SNPs which are obvious significant correlated to the target traits under a loose condition of P-value. After that, we choose the selected SNPs in bigger samples and genotyping. At last, we combine the results of two stages and do the statistics. Compared with the One-stage design. Two or multiple stage design is more economical and efficient strategy.

Now I am going to illustrate a specific design which make use of the information before

The case-control design:

As I mentioned before, one of the most common analysis is case-control analysis, so we start with case-control design first. If there are sufficient numbers of case and control participants, this study design will have a short duration and lower cost than alternative designs and is well suited for investigations of both rare and common diseases. But sometimes, the case-control design can be subject to substantial biases and spurious associations unless both the case and control subjects are correctly classified.<sup>10</sup> Therefore, selection of participants is a key component in the reliability of the case-control study. Samples from control group should be derived from the same population as the group with traits. It's better for them to share the same risk of develop traits. Control subjects may have an early undetected disease. In common diseases such as stroke, efforts must be made to ensure that controls are disease free. So-called super controls, people with high risk but without evidence of disease, have been used in a number of GWA studies.<sup>11</sup> There are few points that you should know: 1). Establish principles for design, such as description and comparison of important phenotypic traits for two groups. 2). Provide a table of comparing two groups, which can make it easier to assess two groups. 3). The limitation

of the case-control design is that the temporal relationship between SNPs and disease state is not captured, although this is generally less of a concern for genetic studies, where alleles are inherited at birth and do not change. For example, a stroke may not be recorded; therefore, the reported age of onset is inaccurate.<sup>12</sup> 4). Many case-control GWA studies make use of a multiple-stage design to increase efficiency and powerfulness.

Besides, there are many other popular designs such as “The Cohort Design” and “The Trio Design”, you can easily find them with other resources.