Miniature Genome Wide Association Study for NC DNA Day

This activity is designed to simulate a (miniature) Genome Wide Association Study with the students as the study participants. Each student will be assigned a phenotype (blue or white) and genotypes at three different variants on the human genome. I thought it would be fun for the phenotype we're testing to be a superpower (e.g. invisibility) rather than a disease, but you can of course use whatever you like. Blue is the affected phenotype (superpower or disease) and white is the normal phenotype.

Materials:

20 slips of white paper listing genotypes 20 slips of blue paper listing genotypes DNAdayGWAS.xlsx (graphs your results for you)

Instructions:

- 1. Count the number of students in your class. Pick out an equal number of slips of paper; half white and half blue. For example, if there are 30 students pick 15 white slips and 15 blue slips.
- 2. Open DNAdayGWAS.xlsx
- 3. Pass out the slips of paper, one slip to each student.
- 4. Read this to the students: "We're going to do our own miniature Genome Wide Association Study, using you as the study participants. A new mutant trait has emerged in humans, invisibility, and we're trying to figure out which variants are responsible. We have three variants that we suspect of being correlated and are going to test: chromosome 1 position 1000, chromosome 2 position 1000 and chromosome 3 position 1000. Each of you has a piece of paper indicating your genotypes at those positions. If you have a white slip of paper, you're normal (no invisibility, sorry!). If you have a blue slip of paper, you have the trait of invisibility. We're going to test each variant and decide whether it is correlated to invisibility or not."
- 5. Starting with chr1:1000, ask students with AA genotypes at chr:1000 to raise their slips of paper. Count the number of white slips and enter that number into the corresponding cell (C2) in DNAdayGWAS.xlsx. Do the same for the number of chr1:1000 AA genotypes that are blue (cell D2). Then ask students with an AG genotype at chr1:1000 to raise their papers, counting white then blue. Then GG genotypes.
- 6. Once you finish all the genotypes for chr1:1000, the plot* for chr1:1000 correlation should be finished with a best fit line and its equation shown. Ask the students if they think the variant at chr1:1000 is correlated to the trait. Generally, if the plot has a steep slope, then this suggests that it's correlated. A slope close to 0 suggests no correlation. What exactly counts as a "steep slope?" Well, this depends on your interpretation, your study size, what you think is reasonable etc. You can discuss this with the students. Note that here, the steepest slope you can really get is .5.
- 7. Repeat steps 5 and 6 for chr2:1000 and chr3:1000.

Results:

Your results may vary if you have a class smaller than 40 due to sampling from the blue and white slips. Generally, you should get a slope close to 0 for chr1:1000, a slope close to .5 for chr2:1000 and a slope close to -.2 for chr3:1000. So chr1:1000 is definitely not correlated while chr2:1000 definitely is. chr3:1000 seems to be correlated somewhat, just not as strongly as chr2:1000. The slope for chr3:1000 is negative simply because of the way the genotypes are arranged on the x-axis. If you flipped their ordering you'd get a positive slope. It's the degree of the slope that matters.

*Plot clarification: The "trait average" on the y-axis is the average value for each genotype given the number of students that were blue and the number that were white, where blue has a value of 1 and white has a value of 0. Therefore, if a genotype has a trait average of 1, that means all the individuals with that genotype had the trait (were blue). A value of 0 means all the individuals were white. Trait averages can thus range between 0 and 1.