

NucleOlap: Finding Candidate Loci Responsible for Recessive Traits in Nuclear Families using SNP Microarray Data

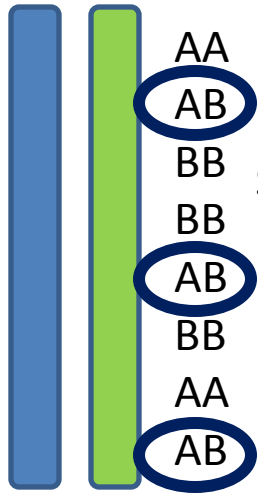
Chromosoft, LLC.

Identity by Descent (IBD) Analysis

- Two Parents with any number of offspring can be analyzed (more than one child is needed to begin narrowing down the candidate regions).
- Parents are used to determine each child's haplotypes (phase is identified by informative SNPs).
- Progeny haplotypes are compared to one another in order to determine the specific regions where alleles are shared (or not shared).

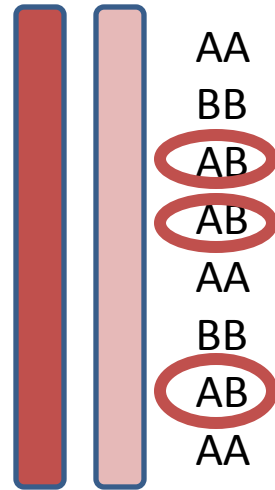
Determining Each Child's Parental Haplotypes

Father



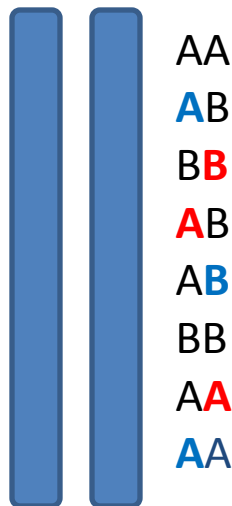
SNPs that are informative are circled for each parent.

Mother



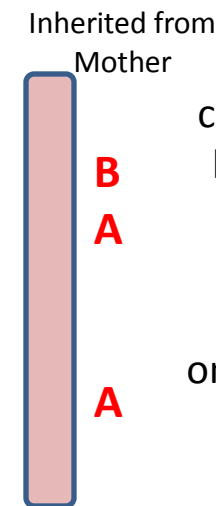
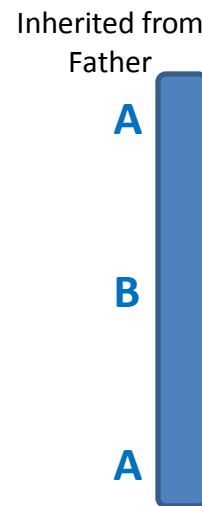
1. Identify informative SNPs
 1. These are SNPs where one parent is heterozygous while the other parent is homozygous.
 2. For these SNPs, it is possible to determine the haplotype inherited by the child for the parent that has the heterozygous SNP.
2. Determine the SNPs inherited by the child from each parent.

Child



The child's genotype is shown with informative SNPs bolded when it is clear that they are inherited from a particular parent.

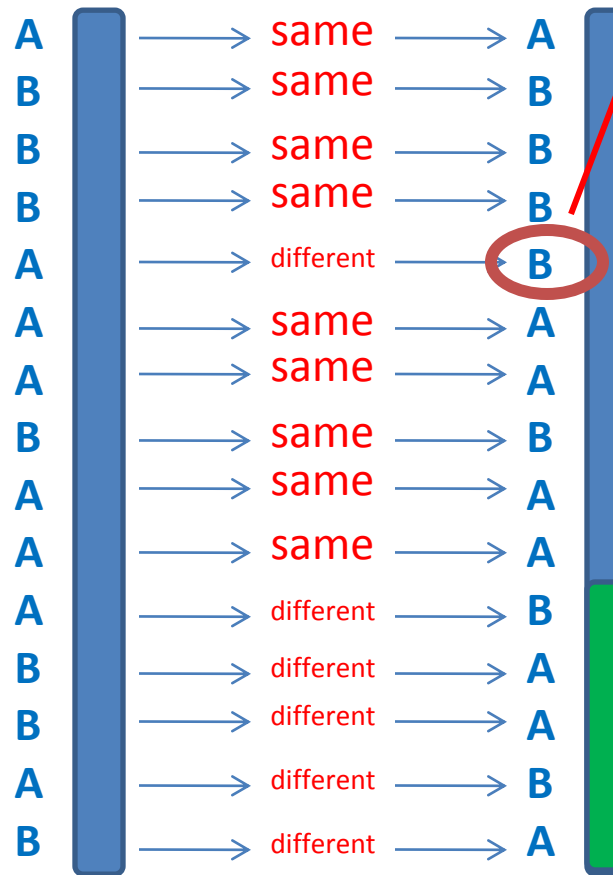
Child



Phased child chromosomes can be inferred from informative inheritance information (we only care about the informative markers).

Determine Recombinations and Eliminate Miscalls

- Phased progeny chromosomes are compared to one another to determine where alleles are shared. Ex. Compare two children with haplotypes from father.
- All children are compared to “child 1” and it is determined if they have the same or different haplotypes at each informative position.
- Miscalls are eliminated if there is not strong evidence for a recombination (i.e. if only one SNP in a large region differs, then we assume that SNP was a miscall, and do not treat it as a crossover event).



This one “different” SNP exists within a region of SNPs that are the same b/t the two children. For this reason, it is not treated as a crossover, but as a mis-called allele from the microarray data, and the children are still assigned the same haplotype in the region.

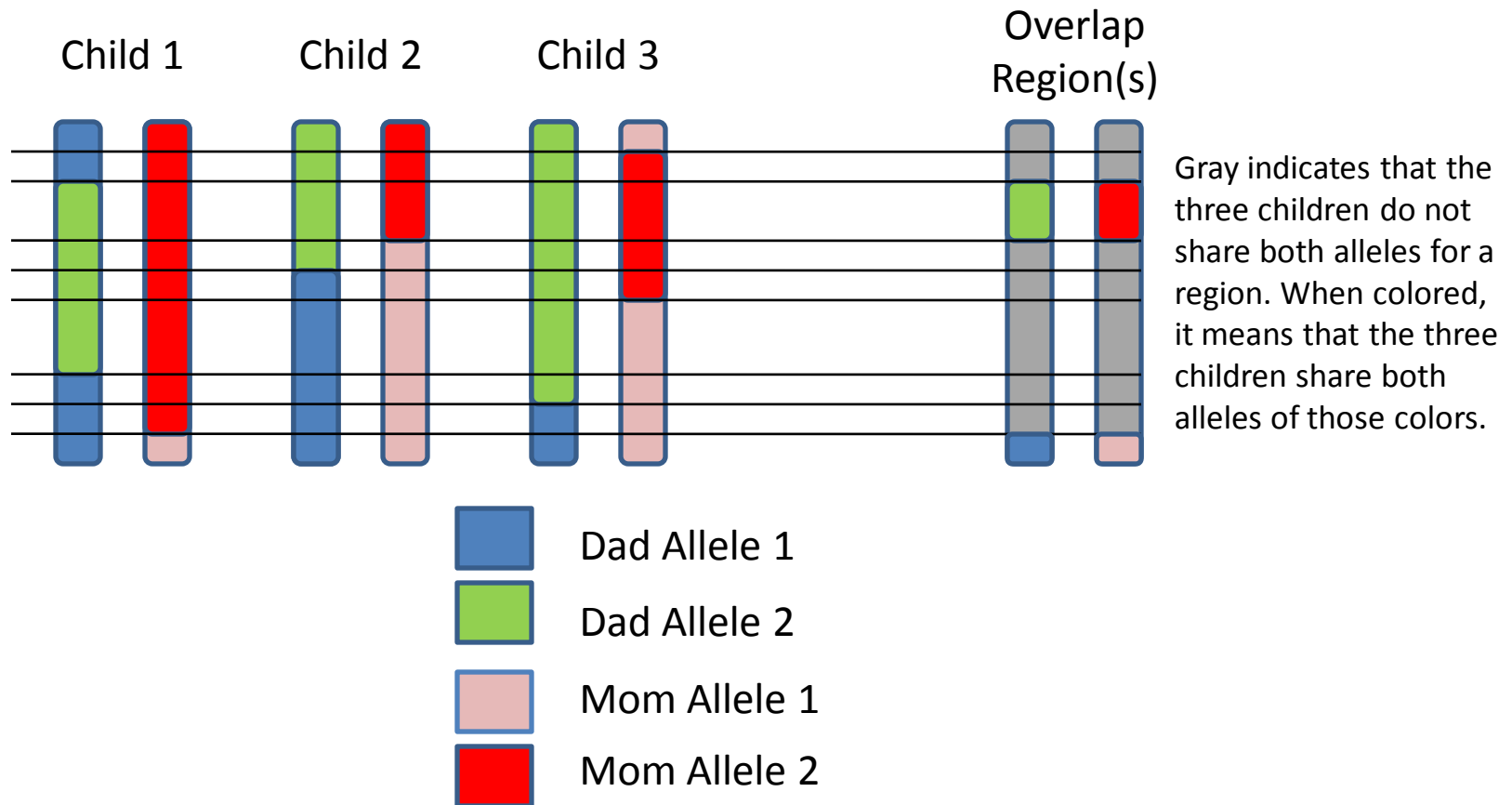
This series of “different” informative SNPs is evidence that a recombination event occurred, and the children share two different haplotypes in this region.

Affected vs. Unaffected Children

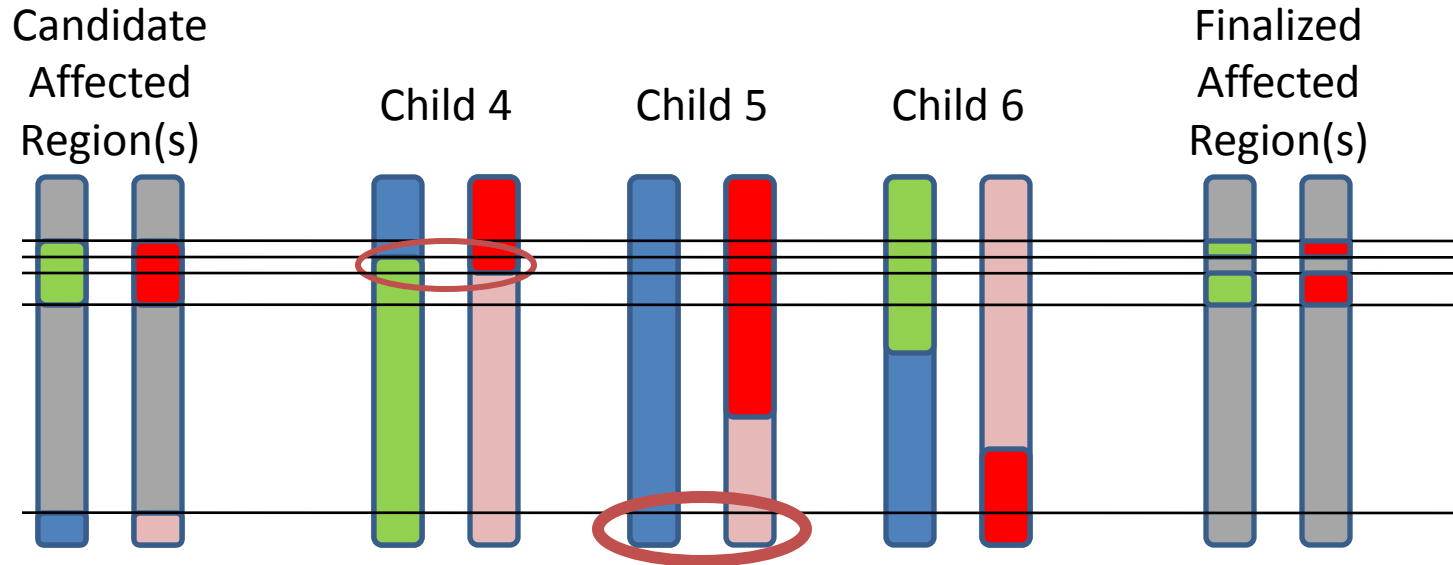
- For family data, children can be identified as affected or unaffected with a particular trait.
- This program assumes that the trait is caused by a recessive mutation, meaning that two copies are necessary for a child to be considered affected.
- Ex. Family of 6: Three affected children, Three unaffected children

Determine Overlap Between Affected Children

- Once phase has been determined for all children, we analyzed the affected children to determine regions where they share both alleles.



Finally, Rule Out Regions where an Unaffected Child Shares Both Alleles



- Dad Allele 1
- Dad Allele 2
- Mom Allele 1
- Mom Allele 2

NOTE: If any one unaffected child shares both alleles with a candidate affected region, that region is no longer a candidate for causing the recessive trait.