PedSplit: pedigree management for stratified analysis

Lanktree MB\textsuperscript{1,*}, VanderBeek L\textsuperscript{1}, Macciardi FM\textsuperscript{1,2}, Kennedy JL\textsuperscript{1}

\textsuperscript{1} Neurogenetics Section, Centre for Addiction and Mental Health, Department of Psychiatry, University of Toronto, 250 College Street, Toronto M5T 1R8, Canada.
\textsuperscript{2} Department of Human Genetics, University of Milan, Italy.

*To whom correspondence should be addressed
ABSTRACT

Summary: PedSplit facilitates pedigree management for gene-gene interaction and sex specific tests for increased homogeneity within subgroups. PedSplit also provides a simple approach for calculating Haplotype Relative Risk (HRR) and generating internal “controls”.

Availability: Executables, C++ source code and documentation for PedSplit can be downloaded from http://www.pharmacogenetics.ca in the links and software section.

Contact: Matthew_Lanktree@camh.net; James_Kennedy@camh.net
The identification of genetic risk factors in common complex disease is a major issue in contemporary research, searching for methods that are powerful and able to address the complexity of these diseases. It has now been established that multiple genetic and non-genetic risk factors play important roles in diseases such as heart disease, cancer, and mental illness. The issues of genetic heterogeneity and epistatic or gene-gene interactions are obstacles that must be addressed (Kennedy et al., 2003). Secondly, “Cut and Paste” errors are an often-overlooked source of error in genetic association studies. PedSplit is a pedigree management utility that provides a fast, error-free method for Haplotype Relative Risk (HRR), sex stratification and Conditional Transmission Disequilibrium Test (CTDT).

**Haplotype Relative Risk (HRR)**

PedSplit generates files containing the probands as cases and the non-transmitted alleles as internal “controls”. The allele counts are provided as output for relative risk calculation. HRR can be more powerful than TDT (Speilman et al., 1993) because more information is included in variance calculations, but TDT may be more powerful in some other situations (Schaid and Sommer, 1994). See Falk and Rubinstein (1987), Ott (1989), and Terwilliger and Ott (1992) for justification.

**Sex Stratification**

The stratification of a data set by the proband’s sex can remove heterogeneity in the case of a sexually dimorphic phenotype (as in OCD; Karayiorgou et al., 1999; Camarena et al., 2001). PedSplit creates two files, one containing families with male probands, the other with female probands, removing the chance of “cut and paste” errors.

**Conditional Transmission Disequilibrium Test (CTDT)**

The invention of high-throughput genotyping systems is revolutionizing the methodology of genetic investigations. Linkage analysis has been very successful at mapping genes of large effect, but the prospect of genome wide association studies will increase the power to detect genes with small effect (Risch et al., 1996). However, simply adding the small effects of many contributing loci may prove to be inadequate for determining the relationship between gene and phenotype in complex disorders. Methods for determining epistatic, gene-gene interactions will need to be developed. PedSplit implements CTDT as described by Culverhouse et al. (2002), an elementary multilocus model for investigating purely epistatic interactions. For example, we examine two loci (alleles A,a and B,b) with no biased transmissions when tested individually. However, when we examine the group of families in which A is transmitted, we note that B is over-transmitted. Further, in the group of families in which a is transmitted, b is over-transmitted. Thus, when the two groups are combined, the group with over-transmitted B and the other with over-transmitted b “cancel” each other out, showing a “neutral” effect, but when the two loci are examined simultaneously a correlation with the phenotype is found. The CTDT command in PedSplit stratifies the sample of n trios by proband genotype at one loci. Haplotype and TDT analysis should then be performed on each stratum of the data. The p-values for each stratum are combined using Fisher’s (1932) statistic ($S = -2 \sum_{i=1}^{m} \ln(p_i)$, where $p_i$ is the TDT p-value corresponding to the $i$th stratum of the data). Under the null hypothesis, $S$ has a $\chi^2$ distribution with $2m$ degrees of freedom, where $m$ is the number of strata (Culverhouse et al., 2002). In conclusion,
PedSplit CTDT can find associations between multilocus genotypes and phenotypes, even when individual loci show no main effect.

References


