

2013年9月30日

遺伝統計解析仕様書：小家系を対象とした簡易的なパラメトリック連鎖解析

株式会社スタージェン 遺伝統計解析事業部

1 解析目的

GWAS チップで観測された SNP 遺伝型データを用いて、家系情報を利用した表現型関連領域の絞り込みを行う。

2 解析方法

パラメトリック連鎖解析を行う。特に本解析では複数の座位との関連性を議論する多点連鎖解析（多点解析とよぶ）を行う。

2.1 表現型関連座位への仮定

2.1.1 頻度

メンデル型形質を仮定するため低頻度アレル頻度を 0.0005 と設定する。

2.1.2 浸透率

遺伝型に対する表現型のとりうる確率（浸透率）を複数設定する。表現型関連座位に存在するアレルを a, A とし、アレル a を関連アレルとする。このとき遺伝型 aa, aA, AA に対して次の浸透率を設定し解析する。

解析番号	浸透率
解析 1	a/a : 0.99, a/A : 0.99, A/A : 0.01
解析 2	a/a : 0.99, a/A : 0.5, A/A : 0.01
解析 3	a/a : 0.99, a/A : 0.01, A/A : 0.01

2.2 解析対象 SNP の選択

パラメトリック連鎖解析は家系情報を用いて複数の観測 SNP 座位と表現型関連座位との間の組換え割合を推定することによって関連領域を絞り込む方法である。そのため一般の GWAS チップで観測された SNP 座位をすべて利用することは冗長な情報になるばかりか誤りの原因となる。本解析では次の条件を満たす SNP 座位を選択して解析を行う。

- (1) 連鎖不平衡状態の強い座位は代表する SNP (tagSNP) を解析対象とする。基準は連鎖不平衡係数 r-square 値が 0.8 以上の領域から 1 つの tagSNP を選択する。
- (2) 1 回の交叉が期待される SNP 間の距離を 1M (モルガン、1M=10⁸ bp 程度) とすると、経験上 1,000 SNP あたり 1 つ程度の SNP 座位を選択する。

2.3 表現型関連領域の絞り込み

関連性の強さを表す LOD スコアを用いて表現型との関連領域を絞り込む。通常は LOD スコアが 3 以上の領域を表現型関連領域と推測するが多いが、小家系の場合は高い LOD スコアを得ることが難しい。

本解析は小家系を想定しているため、原則 LOD スコアが 0 以下（周辺の SNP と表現型関連座位は近くにないという説の方が有用の場合）の領域を除外することで関連領域を絞り込む。

3 ソフトウェア

パラメトリック連鎖解析は MERLIN (Abecasis et al. 2002) を利用する。

4 出力結果

4.1 LOD スコア出力図

各染色体ごとに LOD スコアのグラフを出力する。

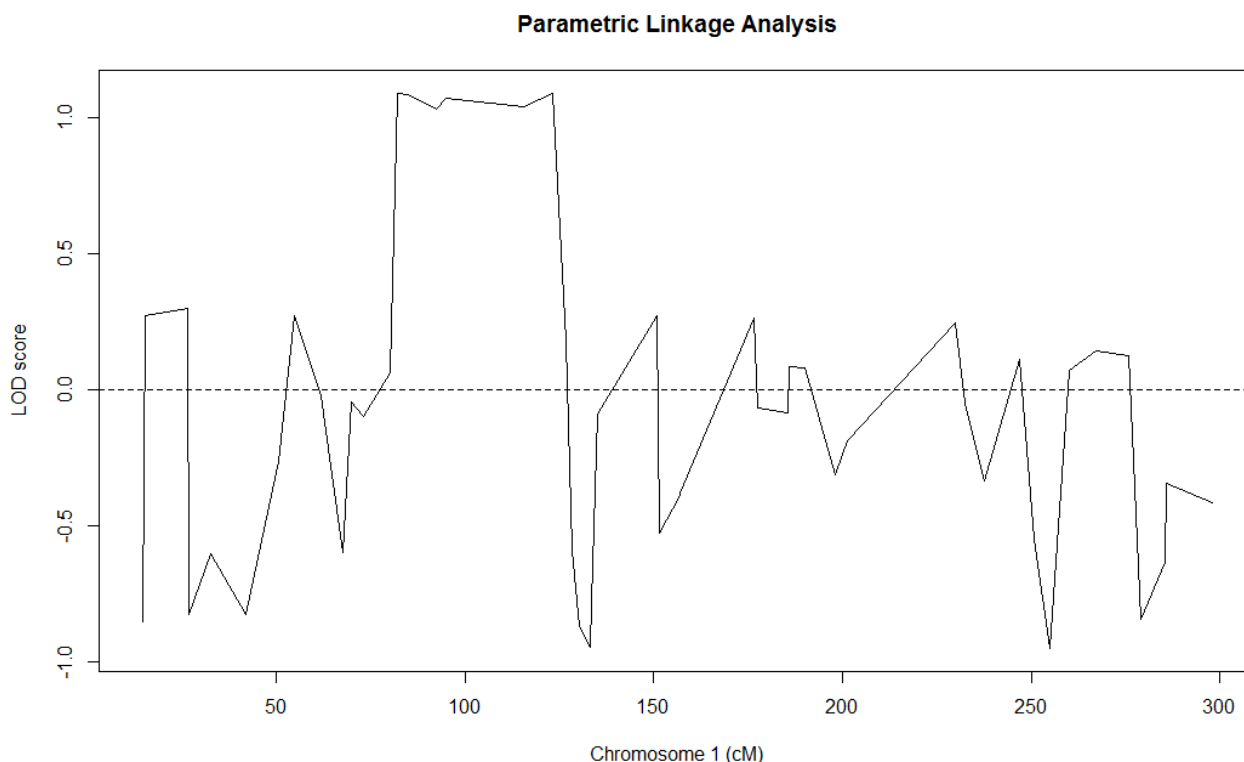


図 1 : LOD スコア出力例

4.2 IGV viewer 対応の BED ファイル

LOD スコアが 0 以上となる領域を候補領域とし、それら候補領域を表した BED ファイルを作成する。

5 参考文献

- [1] Abecasis GR, Cherny, SS, Cookson, WO, Cardon, LR (2002) MERLIN--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97-101.
- [2] Ott J. (1999) Analysis of Human Genetic Linkage 3rd edition. Johns Hopkins University Press. London, UK.