LDBlockShow 使用手册

Visualizing linkage disequilibrium and haplotype blocks based on variant call format files 基于 VCF 文件快速生成连锁不平衡和单体型图

Version 1.35

2020-09-07

hewm2008@gmail.com / hewm2008@qq.com

目录

LDBlockShow 使用手册	1
1.简介	1
2. 下载与安装	1
2.1 下载网址	1
2.2 预先安装	1
2.3 安装	1
3. 参数说明	2
3.1 LDBlockShow	2
3.1.1 主要参数	2
3.1.2 其他参数	3
3.2 ShowLDSVG	4
3.2.1 简要参数	4
3.2.2 详细参数	4
3.3 输出文件	6
4.实例	7
4.1 实例 1 Heatmap + default block generated by PLINK	7
4.2 实例 2: Heatmap + block + GWAS	8
4.3 实例 3: Heatmap + block + GWAS + Annotation	9
4.4 实例 4: Heatmap + Annotation + GWAS plot similar to Locuszoom	10
5. 优势	11
6.常见问题	12
6.1 LDBlockShow 计算 LD 参数的方法	12
6.2 除了 GWAS 分析结果外,可以输入其他统计结果吗?	13
6.3 得到的不是全三角图?	13
6.4 从 SVG 转 PNG 报错,转换失败?	13
6.5 如何显示 SNPs 的名字?	13
6.6 如何改变热图的颜色?	14
6.7 热图看起来跟 R 包 LDheatmap 的结果不太一样?	14
6.8 想要移动热图的位置?	14

1.简介

LDBlockShow 主要用于基于 VCF 文件快速产生连锁不平衡(LD)热图(热图内同时显示单体型范围)。比起其他类似工具,LDBlockShow 所用时间更短,所需内存更小。LDBlockShow 可联合 GWAS 或者其他统计结果以及基因组注释文件一起联合作图。而且 LDBlockShow 支持子群体分析。

2. 下载与安装

2.1 下载网址

https://github.com/BGI-shenzhen/LDBlockShow/

2.2 预先安装

LDblockshow 适用于 Linux/Unix/macOS 系统。在安装之前,请先安装以下使用条件:

1) g++: 推荐 g++ with --std=c++11 > 4.8+ (https://gcc.gnu.org/)

2) zlib:推荐 zlib > 1.2.3 (https://zlib.net/)

3) Perl: Perl 内的模块 SVG.pm (https://metacpan.org/release/SVG)。可以预先安装。尽管我们安装包已经内置了 SVG.pm 包了

2.3 安装

使用者可采用以下三种方式来安装:

1)

git clone https://github.com/BGI-shenzhen/LDBlockShow.git chmod 755 configure; ./configure; make;

mv LDBlockShow bin/; # [rm *.o]

2)

tar -zxvf LDBlockShowXXX.tar.gz cd LDBlockShowXXX;

cd src;

sh make.sh # or in Linux: make; make clean

../bin/LDBlockShow

3) 对于 Linux/Unix 系统用户,如果编译失败,可向我们索取静态编译好的版本,可解压缩后直接运行。联系方式: hewm2008@gmail.com 或 hewm2008@gq.com。

Note: 如果 link 失败,可尝试重新安装 zip 库 (https://zlib.net/).

Note: 对于 Mac 系统用户,如果 plink 运行失败,请重新下载 Mac 系统版本的 plink (https://www.cog-genomics.org/plink2/),然后放到"LDBlockShowXXX/bin"文件夹下。

3. 参数说明

3.1 LDBlockShow

3.1.1 主要参数

[heweiming@cngb-olog	in-25 bin]\$:/bin/LDBlockShow			
Usage: LDBlockShov	v -InVCF	<in.vcf.gz> -OutPut <outprefix> -Region chr1:10000-20000</outprefix></in.vcf.gz>			
1.1/05					
-InVCF	<str></str>	Input SNP VCF Format			
-OutPut	<str></str>	OutPut File of LD Blocks			
-Region	<str></str>	In One Region to show LD info svg Figure			
-SeleVar	<int></int>	Select statistic for deal. 1: D' 2: R^2 3/4:Both [1]			
-SubPop	<str></str>	SubGroup Sample File List [ALLsample]			
-BlockType	<int></int>	method to detect Block [beta] [1]			
		1. Block by PLINK (Gabriel method withed D')			
		2. Solid Spine of LD RR/D' 3. Blockcut with self-defined RR/D'			
		4. FixBlock by input blocks files 5. No Block			
		1. Tixblock by imput blocks files 6. The block			
-InGWAS	<str></str>	InPut GWAS Pvalue File (chr site Pvalue)			
-InGFF	<str></str>	InPut GFF3 file to show Gene CDS and name			
-BlockCut	<float></float>	'Strong LD' cutoff and ratio for BlockType3 [0.85:0.90]			
-FixBlock	<str></str>	Input fixed block region			
-MerMinSNPNum <int></int>		merger color grids when SNPnumber over N[50]			
Morning Train and Menger solor grids when our namber over [10]					
-help		Show more Parameters and help [hewm2008 v1.22]			

-InVCF VCF 格式的输入文件

-OutPut 输出文件路径和文件名前缀 (e.g., /path/pop1)

-Region 产生热图的区域 (格式: chr:start:end)

-SeleVar 使用的 LD 参数 (1: D' 2: R^2; 3 或 4: 3 or 4: 两种 LD 参数都会输出), 默认 1. 当与 GWAS 图联合显示时, 如果用户想要在 GWAS 图种显示最显著 SNP(由

ShowLDSVG 内的—TopSite 指定) 和其他 SNPs 之间的 LD 关系: 使用 1 则热 图和 GWAS 图均使用 D'; 使用 2 则热图和 GWAS 图均使用 R^2; 使用 3 则 热图使用 D'而 GWAS 图使用 R^2; 使用 4 则热图使用 R^2 而 GWAS 图使用 D';

-SubPop 子群体分析的样本名称列表

-BlockType Block 的定义方式。一共有四种,默认的 1 是调用 PLINK ¹生成的,其定义方

式是根据 Gabriel $et\ al.^2$ 发表的文章.2 是 solid spine of LD^3 ,即只考虑倒三角的边缘线上的 LD 情况。用户也可以选择 3 并联合"-BlockCut"命令自定义 r^2 或者 D' 的界值或者选择 4 并联合"-FixBlock"命令提供自己定义

的 block 区域. 如果用户不想显示 block 区域, 可输入参数 5。

-InGWAS 输入与 LD 热图一起画图的统计结果信息(比如 GWAS 的关联结果,其他统

计结果比如 Tajima's D 也可作为输入)文件格式: [chr position Pvalue]

-InGFF 输入 GFF3 格式的文件用于基因组区域注释

-BlockCut 对于类型 3 的 block, 联合该命令定义强 LD 的界值和 block 内强 LD 的 SNPs

所占的比例。默认是0.85:0.9. 也就是,如果在-Sele Var选了D'作为统计量,

那么一个 block 内,两两 SNP 的 D'超过 0.85 的比例为 0.9。

-FixBlock 对于第四种类型的 Block, 用户可以使用此选项提供自定义的 Block 区域。

该文件包含三列,包括染色体、开始位置和结束位置。

-MerMinSNPNum 合并相同色块的最小 SNPs 个数, 默认是 50。关于合并色块, 详见图 1。

-help 显示更多参数

3.1.2 其他参数

[heweiming@cngb-ologin-25 bin]\$./bin/LDBlockShow _h More Help document please see the Manual.pdf file Para [-i] is show for [-InVCF], Para [-o] is show for [-OutPut], Para [-r] is show for [-Region] -InGenotype InPut SNP Genotype Format <str> -InPlink InPut Plink [bed+bim+fam] or [ped+map] file prefix <str> -MAF Min minor allele frequency filter [0.05] <float> -Het <float> Max ratio of het allele filter [0.90] -Miss <float> Max ratio of miss allele filter [0.25] 'Strong LD' cutoff for TagSNP [0.80] -TagSNPCut <float> -OutPng convert svg 2 png file -OutPdf convert svg 2 pdf file

-InGenotype 除了 VCF 格式文件以外,还可以输入 Genotype 格式的文件,格式如下:

-InPlink 支持输入 PLINK 格式的文件,该参数输入 PLINK 格式文件名的前缀

-MAF 过滤最小等位基因频率过低的点 (default ≤0.05)

-Het 过滤杂合度过高的点 (default ≥0.9)

-Miss 过滤分型率失败率过高的点 (default ≥0.25)

-TagSNPCut 挑选 Tag SNPs 的 LD 界值,默认 0.8。

-OutPng 将 SVG 文件转换为 PNG 文件。
-OutPdf 将 SVG 文件转换为 Pdf 文件。

备注: 如果输出 SVG 文件比较小,但用户仍无法打开 SVG 文件,建议使用"-Outpdf"来使用 pdf 文件。在输出的 SVG 文件比较大的情况下,建议使用"-OutPng"转换为 PNG 文件来打 开。

3.2 ShowLDSVG

ShowLDSVG 用于改善 LDBlockShow 得到的图形显示效果(比如修改颜色)。

3.2.1 简要参数

./bin/ShowLDSVG

Options

-InPreFix <s>: InPut Region LD Result Frefix

-OutPut <s>: OutPut svg file result

-help : Show more help with more parameter

-InPreFix 输入文件的前缀名 (i.e., LDBlockShow 的输出文件前缀名)

-OutPut 输出文件前缀名(svg, png 和 pdf 格式的图形将会被输出)

-help 更多参数

3.2.2 详细参数

-InGWAS 输入与 LD 热图一起画图的统计结果信息(比如 GWAS 的关联结果, 其他

统计结果比如 Tajima's D 也可作为输入)文件格式: [chr position Pvalue]

-NoLogP 默认情况下,-InGWAS 输入文件的 P 值将会被-log10 转换,使用该命

令可取消转换。

-TopSite -TopSite 默认值是 GWAS 最显著的点,用户也可以输入 chr:pos 来指定

其他点。使用该命令后,该位点在 GWAS 图中将被显示为菱形。其他

点与该点的 LD 关系将用不同的颜色显示。例图见图 9.

-PointSize 用户可使用该选项来定义 GWAS 点的大小、任意大于 0 的数字均可。

-SpeSNPName 该选项指定的文件可用来输入感兴趣的 SNP 的名字。这些名字会在热

图上方显示。默认输入文件格式是[chr position Name]

-ShowGWASSpeSNP 该选项与-SpeSNPName 一起使用,使用该命令时感兴趣的 SNP 的名字也会在 GWAS 图中。

./bin/ShowLDSVG -h

-InGWAS <s> : InPut GWAS Pvalue File(chr site Pvalue)

-NoLogP : Do not get the log Pvalue-Cutline <s> : show the cut off line of Pvalue

-TopSite <n> : InPut the Special Site as the top site (chr:pos)

-PointSize <n> : set the GWAS point size number

-SpeSNPName <s>: In File for Special SNP Name(chr site Name)

-ShowGWASSpeSNP : show Special SNP Name in GWAS plot with [-SpeSNPName]

-InGFF <s> : InPut GFF3 file to show Gene CDS and name

-NoGeneName : No show Gene name, only show stuct

-crGene <s> :InColor for Gene Stuct [CDS:Intron:UTR:Intergenic]

default: [#e7298a:lightblue:#7570b3:#a6cee3]

-crBegin <s> : In Start Color RGB [255,255,255]

-crMiddle <s> : In Middle Color RGB [240,235,75] -crEnd <s> : In End Color RGB [255,0,0]

-NumGradien <s> :In Number of gradien of color -crTagSNP <s> : Color for TagSNP [31,120,180]

-CrGrid <s> : the color of grid edge [white]
-WidthGrid <s> : the edge-width of gird [1]
-NoGrid : No Show the gird edge

-ShowNum : Show the R^2/D' in the heatmap

-NoShowLDist <n> : NoShow long physical distance pairwise[10000000] -MerMinSNPNum <s> : merger color grids when SNPnumber over N[50]

-OutPng : convert svg 2 png file -OutPdf : convert svg 2 pdf file

-ResizeH : resize image height; Width be resize in ratio [4096]

-InGFF 输入 GFF3 格式的文件用于基因组区域注释; 默认情况下会显示基因名字。

-NoGeneName 使用该命令取消显示基因名。

-crGene 使用该命令定义不同基因组区域的颜色,默认情况下 CDS、内含子, UTR

区和基因间隔区将被分别显示为#e7298a、浅蓝色、#7570b3 和#a6cee3。

用于优化热图颜色的参数:

-crBegin 无 LD 的颜色(R^2/D'=0), 默认白色。

-crMiddle R^2/D'=0.5 的颜色,默认黄色。

-crEnd 完全 LD 的颜色 (R^2/D'=1)默认红色。

-NumGradien 从 crBegin 到 crEnd 的渐变数目。

-crTagSNP 自定义标签 SNP 的颜色。

用于优化热图中方格的参数:

-CrGrid方格边缘的颜色,默认白色-WidthGrid方格边缘的宽度,默认 1

-NoGrid 方格无边缘

-ShowNum 在方格中显示 LD 的数值 (SNP 数目较多, 比如超过 50 时, 不推荐使用此命令).

-NoShowLDist 当两两 SNP 的距离超过一定数值(默认 10M)时,热图中将不会显示这两点之间的 LD 关系。此功能有助于压缩图片大小。用户可自定义此距离。

-MerMinSNPNum 合并相同色块的最小 SNPs 个数, 默认是 50。关于合并色块, 详见图 1。

-OutPng 将 SVG 文件转换为 PNG 文件。
-OutPdf 将 SVG 文件转换为 Pdf 文件。

-ResizeH 可用该参数定义图片的高度来调整分辨率 (默认值 4096)。宽度将会自动等比例调整。

当 SNP 的数目很多,比如大于 100 时,输出的 SVG 文件可能会比较大。ShowLDSVG 会将相邻的颜色一样的方格合并显示,图 1 是一个示意图,可将 SVG 文件从 26k 到 8k。颜色渐变的份数越小,(使用 –NumGradien 定义),压缩效果越明显。-MerMinSNPNum 可用于设置启用合并色块的最小 SNP 数目。

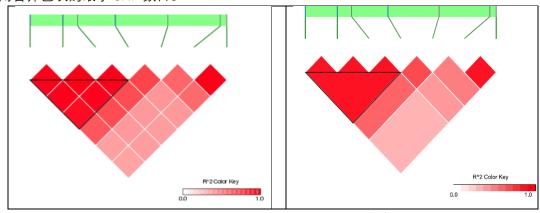


图 1. SNPs 数目较多时候, ShowLDSVG 的压缩效果示例

3.3 输出文件

输出文件	说明
out.site.gz	过滤后剩余的进入计算的 SNPsR [chr site]
out.blocks.gz	Block 文件 [chr start end block_length SNP_number SNPs]
out.TriangleV.gz	区域内成对的 R^2/D'计算结果
out.svg	输出的 SVG 格式图
out.png	输出的 png 格式图
out.pdf	输出的 pdf 格式图

4.实例

所有实例均使用 R^2 作为 LD 参数,但默认值是 D' (不设定-SeleVar 或-SeleVar 1)。

4.1 实例 1 Heatmap + default block generated by PLINK

在 example/Example1 文件夹中, 我们提供了一个用于产生 LD 热图的例子, 其中 block 调用 PLINK 产生。示例命令在 run.sh 文件中:

```
../../bin/LDBlockShow -InVCF Test.vcf.gz -OutPut out -Region chr11:24100000:24200000 -OutPng -SeleVar 2
```

```
sh run.sh
Start Time:
Mon Jun 1 16:30:19 CST 2020
#Detected VCF File is phased file with '|', Read VCF in Phase mode
##Start Region Cal... :Ghir_D11 24100000 24200000; In This Region TotalSNP Number is 7
find blocks...
Start draw... SVG info: SNPNumber :7 , SVG (width,height) = (402.5,297.5)
convert SVG ---> PNG ...
End Time:
Mon Jun 1 16:30:19 CST 2020
```

```
ls
out.blocks.gz out.pdf out.png out.site.gz out.svg out.TriangleV.gz
```

最终结果图如图 2.

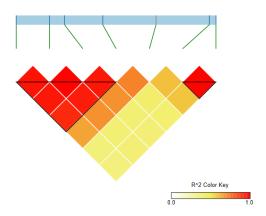


图 2. Example 1 文件夹文件产生的热图

如果选择使用默认的 D'作为 LD 参数, 最终结果如图 3.

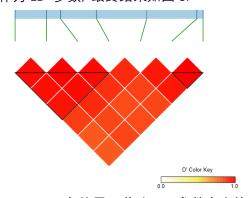


图 3. Example 1 文件用 D'作为 LD 参数产生的结果图

4.2 实例 2: Heatmap + block + GWAS

在 example/Example2 文件夹, 我们提供了用于一起产生热图, block 和 GWAS 统计结果的例子。示例命令在 run.sh 文件中:

../../bin/LDBlockShow - InVCF ../Example 1/Test.vcf.gz - OutPut out - Region chr 11:24100000:24200000 - InGWAS gwas.pvalue - Sele Var 2

结果如图 4 所示。默认情况下, $-\log 10(P \text{ value})$ 大于 7.3 $(P < 5 \times 10^{-8})$ 的点显示为红色。

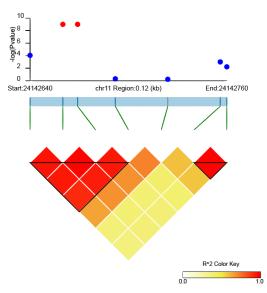


图 4. Example 2 中文件产生的 Heatmap + block + GWAS 图 用户可使用 ShowLDSVG 进一步优化显示效果,示例命令在 run.sh 文件中:

../../bin/ShowLDSVG -InPreFix out -OutPut out -InGWAS gwas.pvalue -Cutline 7 - ShowNum -PointSize 3

优化后的结果如图 5。

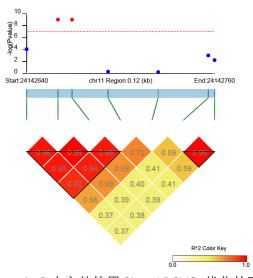


图 5. Example 2 中文件使用 ShowLDSVG 优化的图像结果

4.3 实例 3: Heatmap + block + GWAS + Annotation

在 example/Example3 文件夹中,我们提供了一个同时产生热图、block、GWAS 统计结果和基因组注释结果的例子。示例命令在 run.sh 文件中:

../../bin/LDBlockShow -InVCF ../Example1/Test.vcf.gz -OutPut out -InGWAS gwas.pvalue -InGFF In.gff -Region chr11:24100000:24200000 -SeleVar 2

结果如图 6 所示。

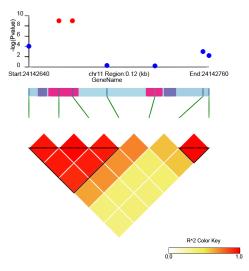


Figure 6. Example 3 中文件产生的 Heatmap + block + GWAS + Annotation 图 用户可使用 ShowLDSVG 进一步优化显示效果,示例命令在 run.sh 文件中:

./././bin/ShowLDSVG -InPreFix out -OutPut out.svg -InGWAS gwas.pvalue -Cutline 7 -InGFF In.gff -crGene yellow:lightblue:pink:orange -showNum 优化后的结果如图 7。

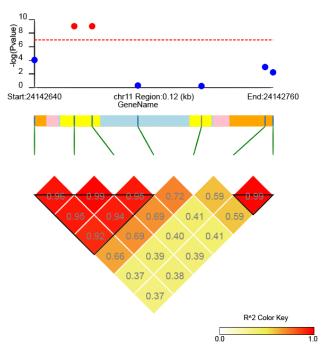


图 7. Example 3 中文件使用 ShowLDSVG 优化的图像结果

用户也可以如下参考命令来显示感兴趣的 SNP 的名字:

../../bin/ShowLDSVG -InPreFix out -OutPut out.svg -InGWAS gwas.pvalue -Cutline 7 -InGFF In.gff -crGene yellow:lightblue:pink:orange -showNum -OutPng -SpeSNPName Spe.snp -ShowGWASSpeSNP 结果图如图 8 所示。

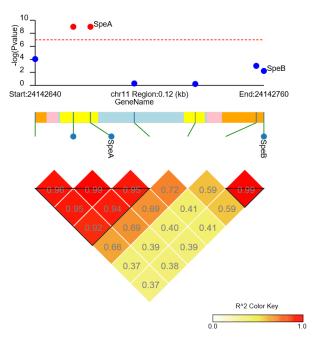


图 8. 优化后显示感兴趣 SNP 名字的结果图

4.4 实例 4: Heatmap + Annotation + GWAS plot similar to Locuszoom

用户可使用-TopSite 来在 GWAS 图中显示最显著位点与其他点的 LD 关系(-TopSite 默认用 GWAS 最显著的位点,用户可以输入 chr:pos 来修改为其他点)使用的 LD 参数可以用-SeleVar 修改.

../../bin/LDBlockShow -InVCF ../Example1/Test.vcf.gz -OutPut out -InGWAS ../Example3/gwas.pvalue -InGFF ../Example3/In.gff -Region chr11:24100000:24200000 -OutPng -SeleVar 3 -TopSite

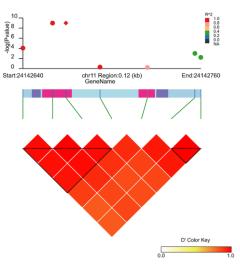


Figure 9. 使用-TopSite 优化后的图可以在 GWAS 图中显示其他点与最显著点之间的 LD。(使用-SeleVar 3, GWAS 图中显示的是 R^2, 热图中使用的是 D', 详见-SeleVar 的解释)

5. 优势

为评价 LDBlockShow 的效果,我们对 LDBlockShow, Haploview³, LDheatmap⁴和 gpart⁵进行了测试。LDBlockShow 计算所得的 r² 和 D¹ 值与其他工具一样。如图 10 所示,与其他工具相比,LDBlockShow 计算时间更短,所用内存更小。

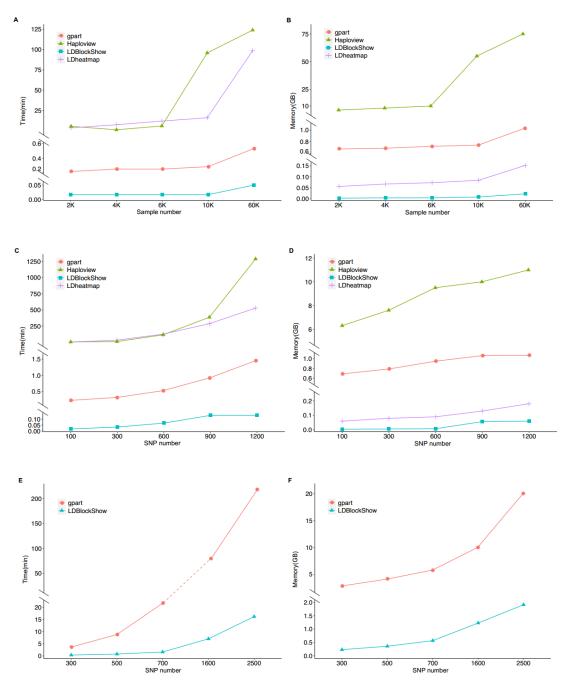


图 10. LDBlockShow、Ldheatmap、Haploview 和 gpart 的计算性能比较。不同方法的 CPU 时间(A)和内存(B) 在 2000 到 60,000 样本中 100 个 SNPs 的计算结果。不同方法的 CPU 时间(C)和内存(D)在 2,000 例样本中, SNPs 从 100 到 1200 个的计算结果。在测试 A-D 数据时,LDBlockShow 和 gpart 的运行时间及内存看起来均在合理范围内。因此,我们测试了它们两个处理较大样本数据的能力。这俩方法的 CPU 时间(E)和内存(F)

在 100,000 例样本中,SNPs 从 300 到 2,500 个的计算结果。计算在 Intel Xeon CPU E5-2630 v4 节点的一个 线程执行。

LDBlockShow 支持同时产生 LD 热图和其他统计结果或者基因组注释结果。此外, LDBlockShow 还支持子群体分析。表 1 是 LDBlockShow 与其他软件的特性比较。

Table 1. LDBlockShow 与其他工具的比较

特性	LDBlockShow	Haploview	LDheatmap	gpart
压缩 VCF 文件作为输入	$\sqrt{}$	×	×	×
非压缩 VCF 文件作为输入	\checkmark	×	×	$\sqrt{}$
子群体分析	\checkmark	×	×	×
<i>输出</i>				
同时产生其他统计结果图	\checkmark	×	×	×
同时产生基因组注释图	\checkmark	×	×	$\sqrt{}$
压缩 SVG 文件	$\sqrt{}$	×	×	×
输出 PNG 文件	\checkmark	\checkmark	×	\checkmark
Block 信息	\checkmark	\checkmark	×	$\sqrt{}$
LD 的统计量	D'/r ²	D'/r ²	r^2	D'/r ²

6.常见问题

6.1 LDBlockShow 计算 LD 参数的方法

与我们之前发表的分析连锁不平衡衰减的文章 6 一样, LDBlockShow 使用以往文献 78 报道的公式计算 2 和 D'. LDBlockShow 与其他工具的计算结果是一样的, 图 11 是与 Ldheatmap 对比的结果图。

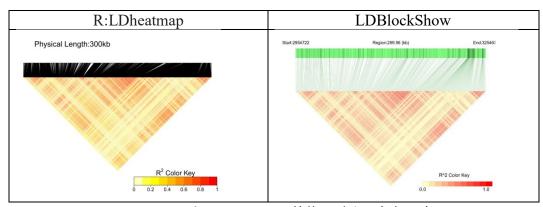


图 11. Ldheatmap 和 LDBlockShow 的热图对比,完全一致。

6.2 除了 GWAS 分析结果外,可以输入其他统计结果吗?

当然可以。-InGWAS 输入文件的第三列可以是任何值。联合使用-NoLogP 命令,第三列的值就不会被 log10 转换。

6.3 得到的不是全三角图?

当区域很大(>1M)时,LDBlockShow 只输出 1M 内的热图。用户可以使用"-NoShowLDist" (默认值 1000000)来改变这个设定。比如使用 "-NoShowLDist 2000000" 会输出 2M 内的热图。图 12 是一个例图。

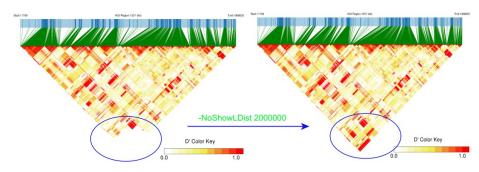


图 12.利用-NoShowLDist 来产生长距离的热图

6.4 从 SVG 转 PNG 报错, 转换失败?

默认情况下, LDBlockShow 会调用用户系统内的"convert"来完成 SVG 到 PNG 的转换。但是有时候,可能由于版本或者文件大小问题,用户系统内的"convert"无法顺利完成转换。这种情况下,用户可使用以下命令来完成转换:

perl LDBlockShow-XXX/bin/svg_kit/svg2xxx.pl xxx.svg -t png --height 50

6.5 如何显示 SNPs 的名字?

用户可联合使用 ShowLDSVG 中的 "-ShowGWASSpeSNP"和"-SpeSNPName"输入的文件来显示感兴趣的 SNP 的名字。具体使用方法可参考例子 3 中的图 8。

6.6 如何改变热图的颜色?

用户可使用 ShowLDSVG 中的 "-crMiddle"和 "-crEnd"来改变热图颜色。比如,如果想要产生黑白色热图,可使用 "-crMiddle 192,192,192 -crEnd 64,64,64".

6.7 热图看起来跟 R 包 LDheatmap 的结果不太一样?

默认情况下,LDBlockShow 用 D'来衡量连锁不平衡。但是 Ldheatmap 只支持输出 r²。若用户想要使用 r²,可使用"-SeleVar 2"来切换。

6.8 想要移动热图的位置?

当显示 SNP 名字的时候,热图的位置会根据 SNP 名字的长度自动调整至最佳。 考虑有的用户可能想要自己微调热图的位置,我们在 ShowLDSVG 内提供了一个参数"-adjHeatmap"。该参数可以接受从-0.25 到 0.25 的数字。其中-0.25 到 0 代表上移,0-0.25 代表下移。比如""-adjHeatmap -0.1" 代表将热图上移幅度 为原画布大小的 10%。

如果有其他需求和建议,欢迎与我联系!加入我们的 QQ 群: 125293663。

Reference

- 1. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
- 2. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225-9 (2002).
- 3. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-5 (2005).
- 4. Shin, J.-H., Blay, S., McNeney, B. & Graham, J. LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. *2006* **16**, 9 (2006).
- 5. Kim, S.A. et al. gpart: human genome partitioning and visualization of high-density SNP

- data by identifying haplotype blocks. *Bioinformatics* 35, 4419-4421 (2019).
- 6. Zhang, C., Dong, S.S., Xu, J.Y., He, W.M. & Yang, T.L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786-1788 (2019).
- 7. Lewontin, R.C. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* **49**, 49-67 (1964).
- 8. Hill, W.G. & Robertson, A. Linkage disequilibrium in finite populations. *Theor Appl Genet* **38**, 226-31 (1968).