

QTL 定位分析报告

合同编号: XX

客户单位: XXX

样本数量: X

报告单位: 北京康普森农业科技有限公司

报告日期: 2023 年 XX 月 XX 日

联系邮箱: agritechnology@kangpusen.com

联系电话: 010-53935082





一、项目概览

本次分析基于 P1、P2 构建的 RIL 群体，共计 XX 个样本，XX 个性状，共鉴定到 XX 个 QTLs。

二、项目基本信息

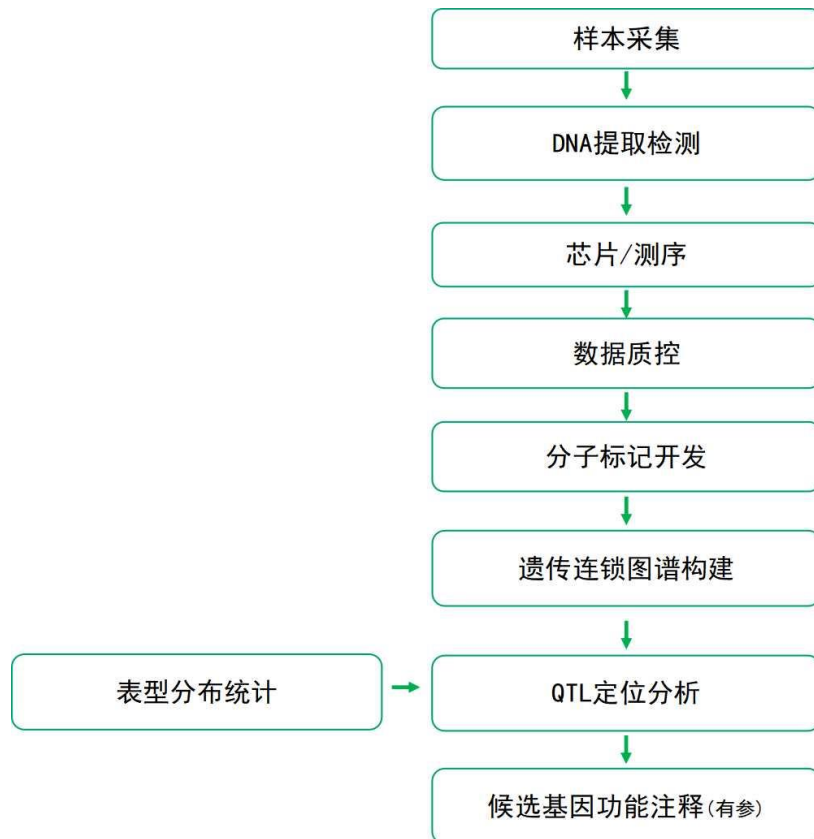
1. 分析内容

尊敬的客户，我们为您提供的报告主要包含以下内容：

文件或文件夹名称		描述
QTL 定位分析报告		对整个项目的综合概述
1 客户提供数据		包含群体信息及表型数据
2 表型统计	2-1 表型数据质控	剔除偏离总体分布个体表型
	2-2 表型分布统计	对研究性状分布的初步了解
3 分子标记开发		标记筛选、位点质控
4 遗传图谱构建	SNP 标记密度分布图	作图标记的密度分布图
	遗传连锁群信息统计	作图标记的遗传位置及物理位置
	连锁图谱可视化	遗传图谱标记分布图
	图谱结果汇总	连锁群遗传图谱信息统计
5 QTL 定位分析		QTL 定位结果汇总
		QTL 定位结果可视化
6 QTL 区间注释		QTL 区间位点注释
		候选基因富集分析



2. 分析流程图



三、分析结果展示及解读

1. 客户提供数据

该结果请见报告附件中“1 客户提供数据”文件夹。该文件夹中是客户提供的用于分析的数据信息，您可对照这个文件夹中的内容进行核实。

2. 表型统计

QTL 定位分析中，一般要求表型数据的随机误差服从正态分布，所以分析前需要对群体表型数据有大致地了解。为方便分析，对目标性状重命名为分析代号，如下表 1 所示，见报告附件中“性状名称对应表.xlsx”。



表 1 性状名称对应表

性状名称	分析代号
性状 1	T1
性状 2	T2
性状 3	T3
性状 N	TN

2.1 表型数据质控

基于作图个体表型数据绘制箱线图，如图 1 所示。依据箱线图，仅将严重偏离总体分布的个体表型替换为 NA。剔除样本明细请见报告附件中“2.表型统计/2-1.表型数据质控/表型质控剔除样本明细.xlsx”。

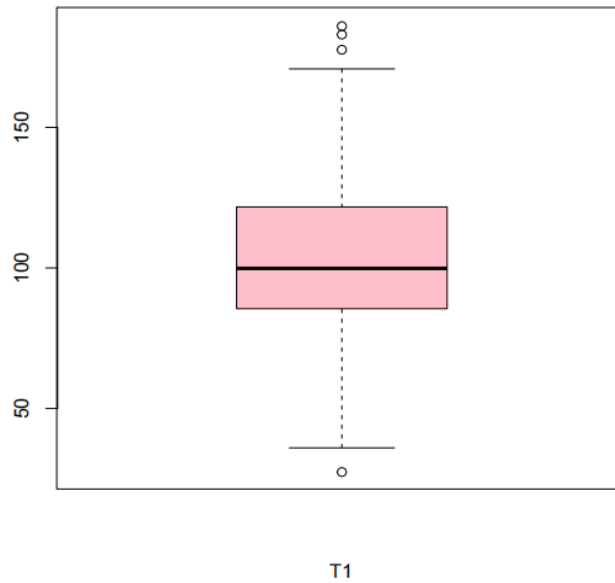


图 1. 性状 N 表型分布箱线图

2.2 表型分布统计

基于作图个体剔除异常值后的表型数据，进行表型分布统计并绘制直方图，如下图 2 所示，该结果请见报告附件中“2.表型分布统计/2-2.表型表型分布统计”。

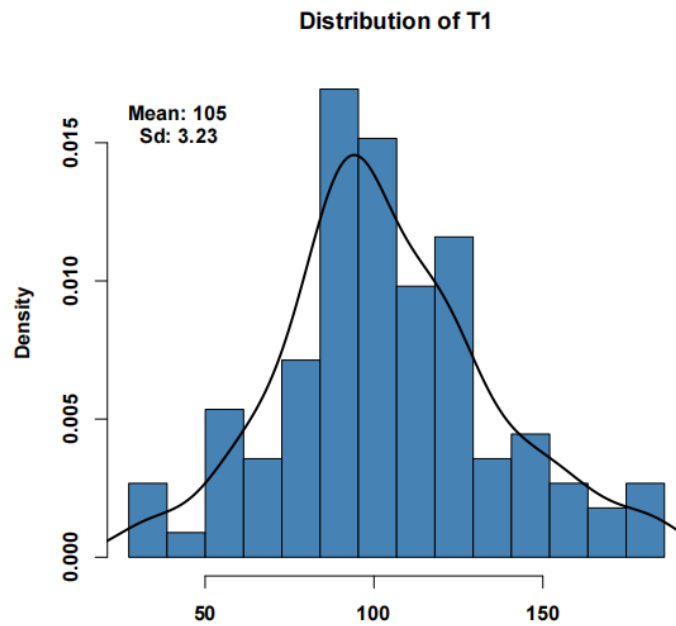


图 2. 性状 N 表型分布直方图

3. 分子标记开发

QTL 定位分析前，首先筛选两个亲本均为纯合基因型且亲本间差异的标记位点，然后对子代群体数据进行基因型质控。对基因型及表型进行一定标准的质控，可以排除异常基因型和异常表型对分析结果产生的影响。

质控条件及质控结果请见报告附件中“3.分子标记开发/个体质控及位点质控统计.xlsx”，质量控制标准见下表 2，质控信息见下表 3，质控后剩余标记的密度分布请见图 3，作图标记请见附件“3.分子标记开发/作图标记.list”。

表 2 质量控制标准

数据	数据质控标准
亲本	亲本间非缺失、纯合且多态的标记位点
子代	a.提取亲本间纯合、多态的位点
	b.SNP 检出率 ≥ 0.90
	c.最小等位基因频率 MAF ≥ 0.05
	d.个体杂合率 ≤ 0.15 (RIL 群体过滤, F_2 群体不过滤)
	e.位点杂合率 ≤ 0.10 (RIL 群体过滤, F_2 群体不过滤)
	f.标记偏离卡方检验 P 值 ≥ 0.00001

表 3 质控信息统计

质量控制标准	数目
标记总数	XXX
亲本间非缺失、纯合且多态标记数	XXX
SNP 检出率 ≥ 0.90 标记数	XXX
最小等位基因频率 $MAF \geq 0.05$ 标记数	XXX
个体杂合率 ≤ 0.15 的个体数	XXX
位点杂合率 ≤ 0.10 标记数	XXX
标记偏离卡方检验 P 值 ≥ 0.00001 的标记数	XXX
用于构建图谱的标记数	XXX



图 3. 筛选用于遗传图谱构建的标记密度分布图

4. 遗传图谱构建

4.1 简介

连锁遗传图又称遗传图谱（Genetic linkage map），是指基因或 DNA 分子标记在染色体上的相对位置与遗传距离，通常以基因或 DNA 片段在染色体交换过程中的重组频率厘摩(cM)表示。1cM 表示两位点在减数分裂中的重组频率为 1%，重组率的值（cM）越高表明两位点之间遗传距离越远，越低表示遗传距离越近。两个基因遗传距离越远，它们之间发生重组事件的概率越高。

遗传图谱常用于数量性状位点 (Quantitative trait loci , QTLs) 定位, 获得与重要农艺性状或抗性性状连锁的基因组片段, 开展基因的图位克隆, 为分子标记辅助选择和育种工作奠定理论基础。

4.2 遗传图谱构建

对筛选后得到的高质量标记, 采用 QTL IciMapping (Version 4.2) 软件进行遗传图谱构建:

- 1) 连锁群划分, 按照标记所在的染色体划分连锁群;
- 2) 采用软件默认参数对每个连锁群标记进行排序;
- 3) 采用 Kosambi 函数计算标记间的遗传距离。

去掉严重不连锁的标记, 最终上图标记 XX 个。每个连锁群上的 SNP 标记数量及每个连锁群的遗传距离总长见下表 4, 遗传图谱上标记分布图请见图 4, 所有上图标记列表请见附件“geneticMap_formatt.csv”。

表 4 遗传图谱信息统计

LinkageGroup	NumMarkers	MapLength.cM	MeanDistance.cM	MaxGap.cM
LG1	26	26.95	1.04	14.99
LG2	170	80.05	3.34	13.15
LG3	188	58.98	0.31	13.35
LG4	28	42.60	1.78	12.10
LG5	24	60.05	3.34	13.15
LG6	123	39.61	0.33	6.16
LG7	157	109.29	0.77	7.29
LG8	46	35.47	0.70	11.29
LG9	32	99.05	3.10	15.52
LG10	369	130.21	0.34	20.69
Total	3682	2043.61	0.56	29.95

注: LinkageGroup: 连锁群编号; NumMarkers: 标记数量; MapLength.cM: 遗传距离长度, 单位是厘摩 (cM); MeanDistance.cM: 平均遗传距离, 单位是厘摩 (cM); MaxGap.cM: 标记间最大遗传距离, 单位是厘摩 (cM)。

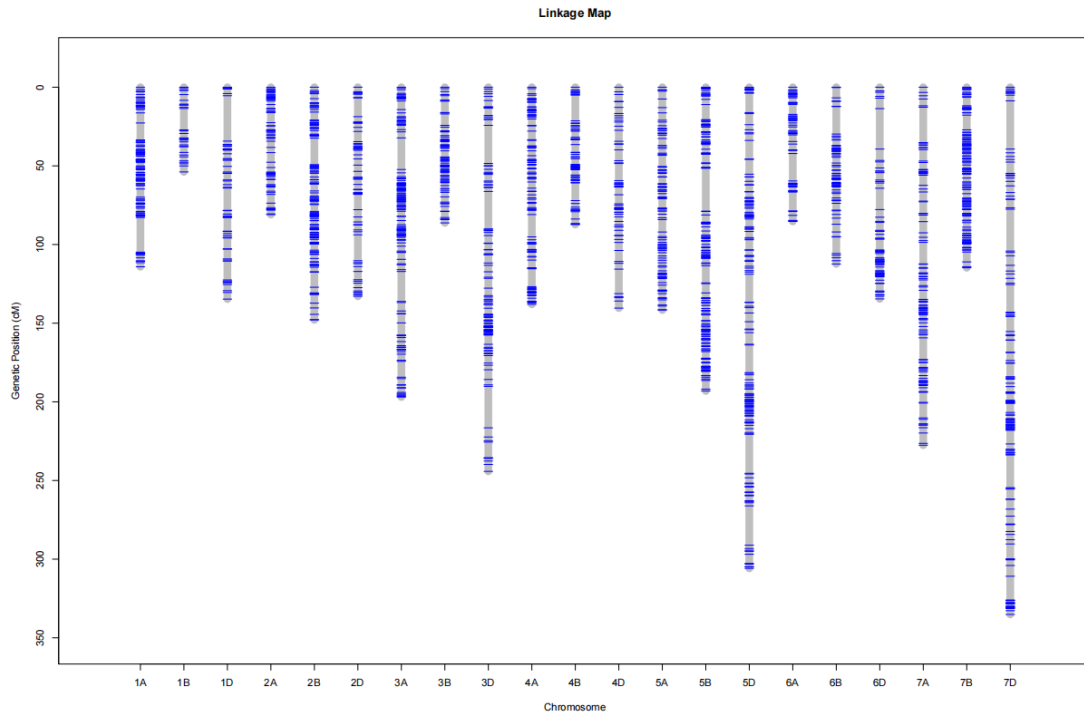


图 4. 遗传图谱标记分布图

5. QTL 定位分析

5.1 简介

植物中大多数重要的性状如株高、产量、抗病、抗逆性等为数量性状，受多基因控制，易受环境影响，呈连续变异。控制数量性状的基因在基因组中的位置称为数量性状基因（Quantitative Trait Locus，QTL）。

QTL 定位是通过数量性状观察值与标记间的关联分析，即当标记与特定性状连锁时，不同标记基因型个体的表型值存在显著差异来确定各个数量性状位点在染色体上的位置、效应。

5.2 分析方法

QTL 常用作图方法包括：单标记或单点分析、区间作图（Interval Mapping, 简称 IM）、复合区间作图法（Composite Interval Mapping, 简称 CIM）、完备区间作图方法(Inclusive Composite Interval Mapping, 简称 ICIM)等。常用的 QTL 作图软件有 QTLIcimapping, MapQTL, WinQTL cartographer, R\qtl 等。



QTL 定位的准确度受群体类型、群体大小、定位方法、分子标记密度等的影响。本项目采用 QTL IciMapping (Version 4.2) 软件进行 QTL 定位，具体分析方法如下：

- 1) 基于完备区间作图法 (ICIM) 进行 QTL 分析，步长设定为 1cM；
- 2) LOD 阈值默认 2.5，可指定 LOD 值；

5.3 QTL 定位结果

本项目共计 XX 个性状，共鉴定到 XX 个 QTLs，QTL 定位结果见下表 5，QTL 定位结果展示见图 5。所有性状 QTL 定位结果见附件“5.QTL 定位分析/*QTL 定位结果.csv”，结果可视化请见报告附件“5.QTL 定位分析/*_QTL 定位结果展示.pdf”。

表 5 QTL 定位结果

Trait	QTL name	LinkageGroup	Position(cM)	Left Marker	Right Marker	LOD	PVE	Add
T1	qT1	1	40	Affx-111002423-19445	Affx-88394661-34511	16.7129	29.9583	25.4477
T2	qT2	5	97	Affx-88590731-34489	Affx-110545577-19541	3.6104	16.5509	-1.7012

注：Trait: 性状；LinkageGroup: 连锁群编号；Position(cM): QTL 区段内 LOD 峰值标记遗传距离，单位是 cM；“Left marker”和“Right marker”: QTL 左侧和右侧边界标记；PVE(%): QTL 解释表型变异的比率；Add(Additive effect):加性效应值。

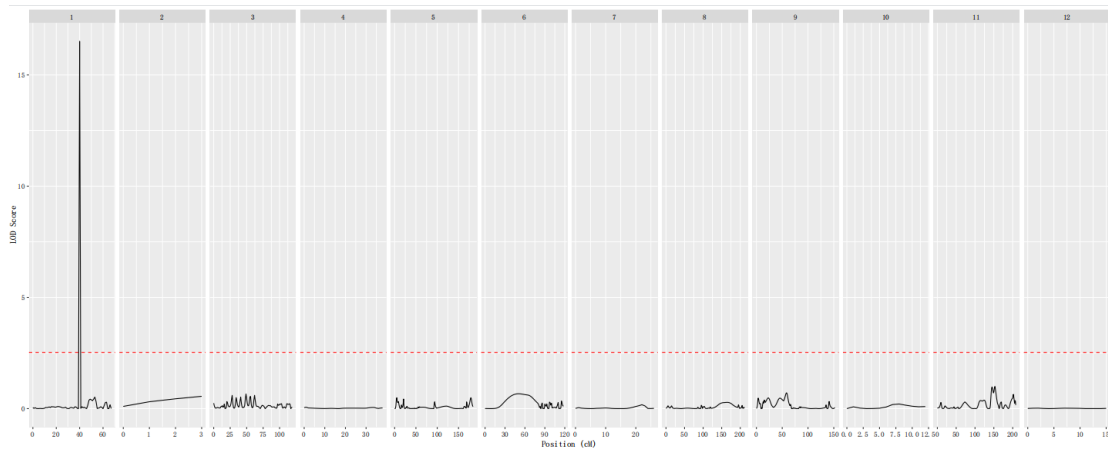


图 5.性状 N QTL 定位结果展示

注：纵坐标为 LOD 值；红色的线为 LOD 阈值线；上横坐标为不同连锁群编号；下横坐标为每条连锁群的遗传距离；



6. QTL 区间注释

6.1 区间位点注释

基于 QTL 定位结果，筛选出与性状关联的 SNP 位点。基于 ENSEMBL 网站上下下载对应物种的参考基因组信息，提取与性状相关 QTL 区间 SNP 位点注释结果，ANNOVAR 注释结果见报告附件中“6.QTL 区间注释/ TN /Anno.variant_function.csv”。

6.2 候选基因富集分析

Gene Ontology（简称 GO）是一个国际化的基因功能分类体系，提供了一套动态更新的标准词汇表（controlled vocabulary）来全面描述生物体中基因和基因产物的属性。GO 总共有三个 ontology，分别描述基因的分子功能（molecular function）、所处的细胞位置（cellular component）、参与的生物过程（biological process）。

和 GO 分析一样，KEGG 通路分析针对的也是候选基因。在生物体内，不同基因相互协调行使其生物学功能，基于 pathway 的分析有助于更进一步了解基因的生物学功能。KEGG 是有关通路的主要公共数据库，通路显著性富集分析以 KEGG 中的通路为单位，应用超几何检验，找出与整个参考基因相比较在候选基因中显著性富集的通路。

基于 ANNOVAR 注释结果，使用 clusterProfiler 包将注释到的区间候选基因根据 GO 和 KEGG 数据库进行基因功能富集分析。GO 富集分析的结果请见报告附件中“6.QTL 区间注释/ TN / GOenrich.csv”，每个 ontology 的点状图请见报告附件中“6.QTL 区间注释/ TN /GOenrich_*_dotplot.pdf”。

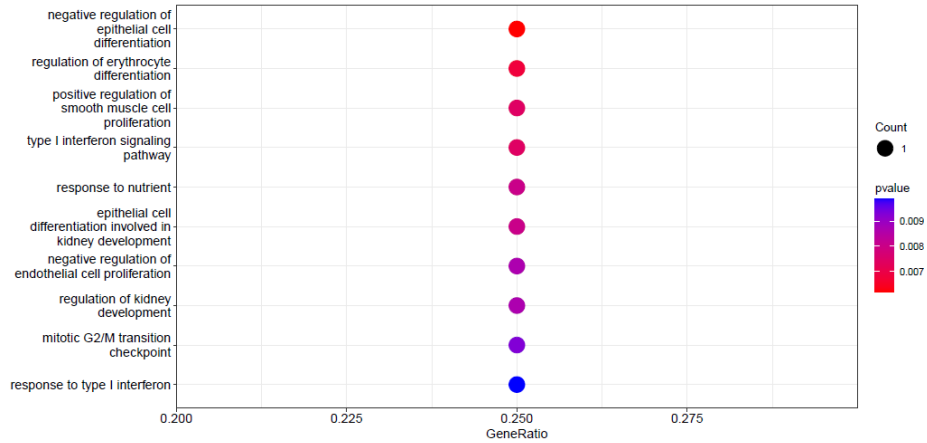


图 6. GO biological process (BP) ontology 富集分析结果点状图

KEGG 通路分析结果请见报告附件中“6.QTL 区间注释/ TN /KEGGenrich.csv”。KEGG 通路分析结果点状图请见报告附件中“6.QTL 区间注释/TN /KEGGenrich_dotplot.pdf”。

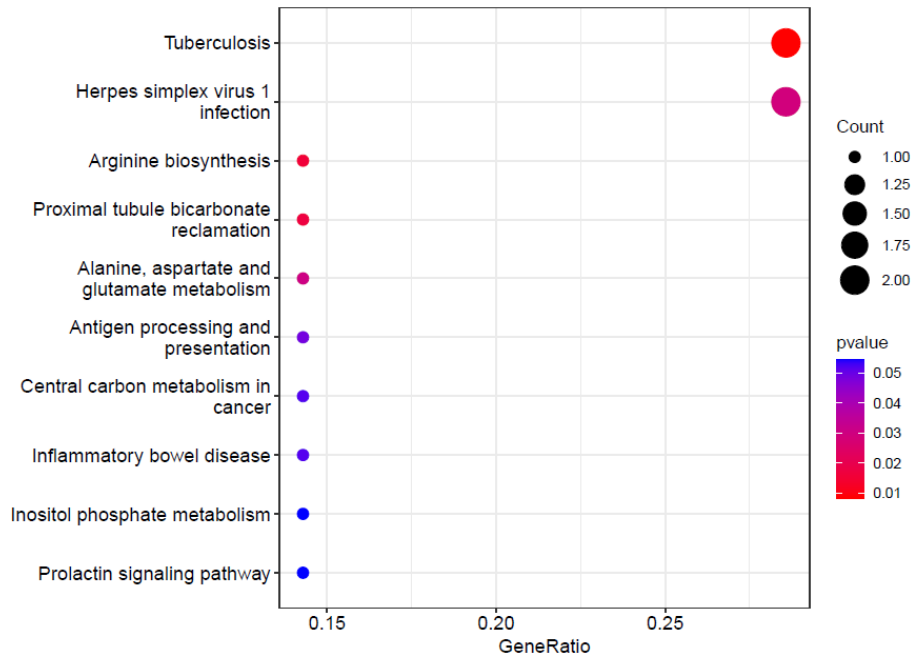


图 7. KEGG 通路分析结果点状图



四、分析结果概述

以上为 QTL 定位分析的所有内容。本项目采用 XX 和 XX 构建的 X 个 X (RIL / F₂/其它) 群体，进行了遗传图谱的构建以及 QTL 定位。

本项目共获得 XX 个 SNP，双亲本均为纯合且亲本间具有多态性的 SNP 为 XX 个。随后对符合作图标记要求的 SNP 进行子代基因分型及标记筛选过滤，共得到 XX 个高质量的 SNP 可用于后续遗传图谱的构建。

本项目利用 QTL IciMapping (Version 4.2) 软件进行遗传图谱的构建以及 QTL 定位，共得到 XX 个连锁群，遗传距离总长约为 XX cM，上图标记总数为 XX 个，标记间的平均遗传距离约为 XX cM。本项目共计 XX 个表型，共鉴定到 XX 个 QTLs。最后，对 QTL 区间位点注释、区间候选基因进行富集分析，详细 QTL 信息及候选基因见报告附件“6.QTL 区间注释”。希望我们的分析结果能对您接下来的研究有所帮助。

五、参考文献

[1]Wang, J, H. Li, L. Zhang, and L. Meng. 2019. Users' Manual of QTL IciMapping. The Quantitative Genetics Group, Institute of Crop Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100081, China.

[2]李慧慧,张鲁燕,王建康.数量性状基因定位研究中若干常见问题的分析与解答[J].作物学报,2010,36(06):918-931.

[3]Wang J-K(王建康). Inclusive composite interval mapping of quantitative trait genes. Acta Agron Sin (作物学报), 2009, 35(2): 239–245 (in Chinese with English abstract)

[4]Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology 16, no. 5 (May 2012): 284–287

[5]Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet., 2007. 81(3): 559–75

[6] Danecek P, et al. The variant call format and VCFtools. Bioinformatics, 2011. 27(15): 2156-2158.