



用eQTL网络研究组织中的调控

专业：生物化学与分子生物学

学号：2017304110114

姓名：贾庆霄

导师：周道绣



背景

在生物界很多性状并不是有单一的基因控制的，而是有多个基因共同调控的，比如说人的身高和很多疾病。虽然人类的基因组序列已经被报道多年，但基因与性状之间的关系却不能很好的说明。通过eQTL网络可以更好的研究数量性状基因怎样调控组织性状。

eQTL（**expression quantitative trait locus**）：数量性状基因组的表达量。

SNP：指在基因组上单个核苷酸的变异,包括置换、颠换、缺失和插入。

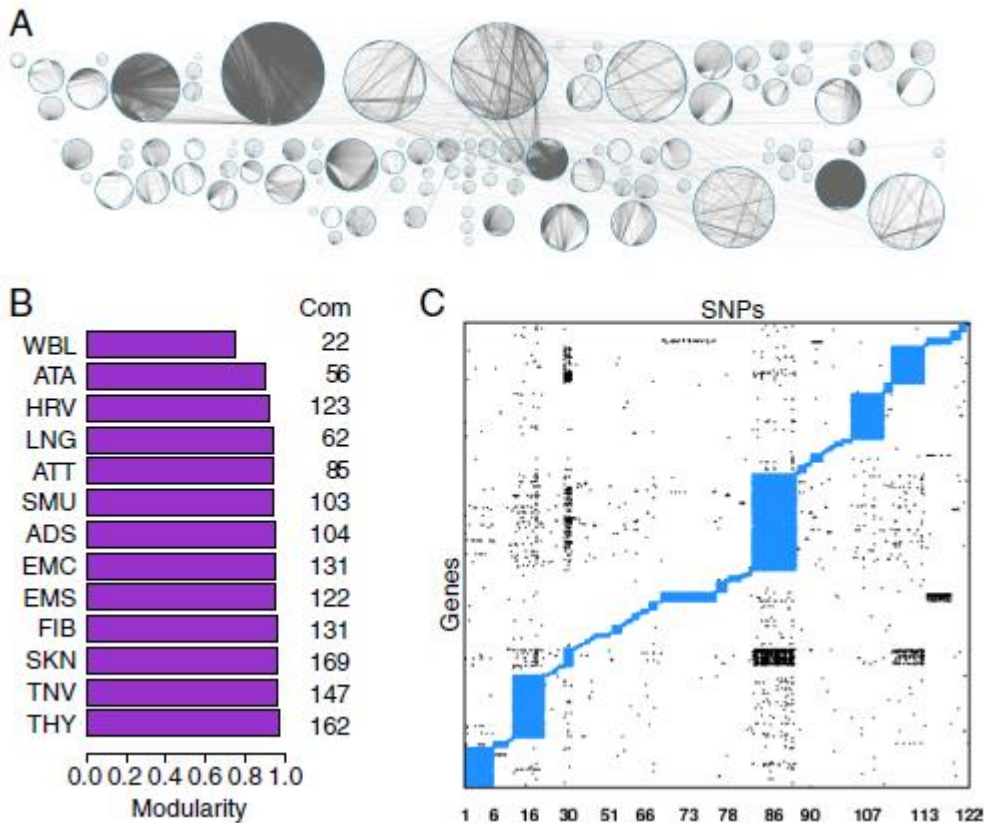


一、eQTL网络高度模块化

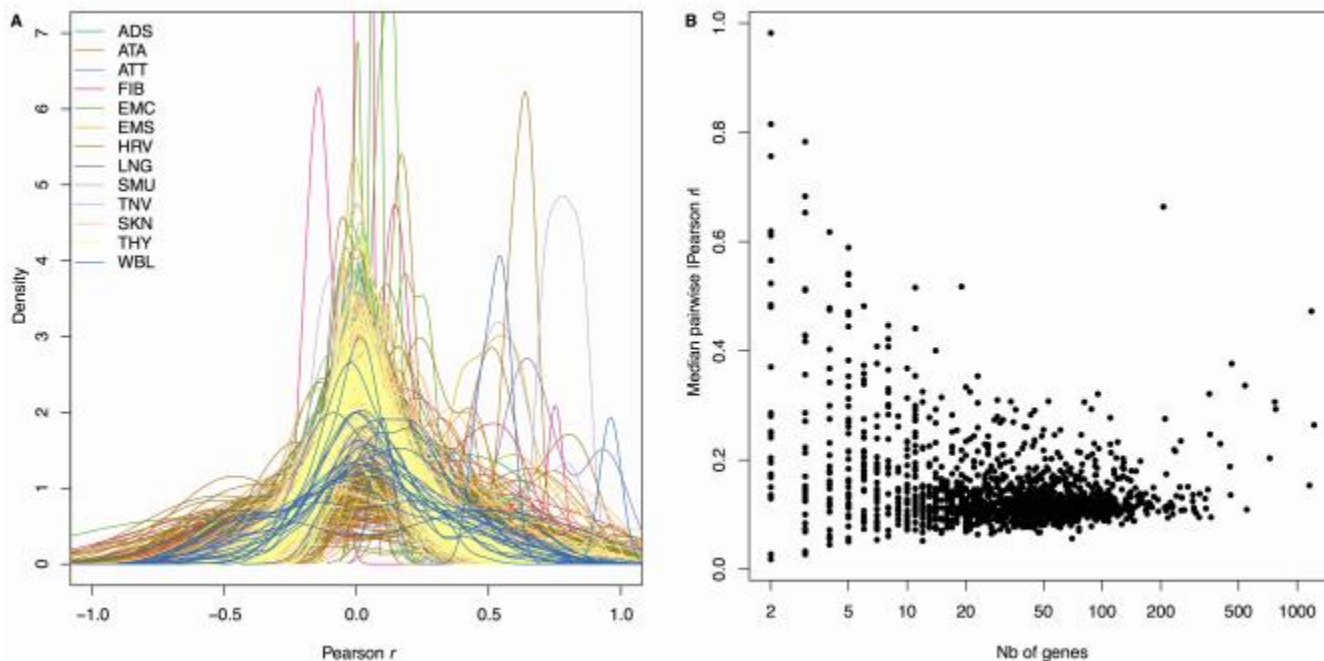
使用GTEx项目提供的数据，进行预处理得到12个组织和1个细胞系，然后对13种不同组织中顺式和反式eQTL构建的eQTL网络进行了系统分析，发现，这些网络的结构特性可以更好的观察跨组织和组织内遗传变异体的调控作用功能。



A图：心脏左心室的eQTL网络



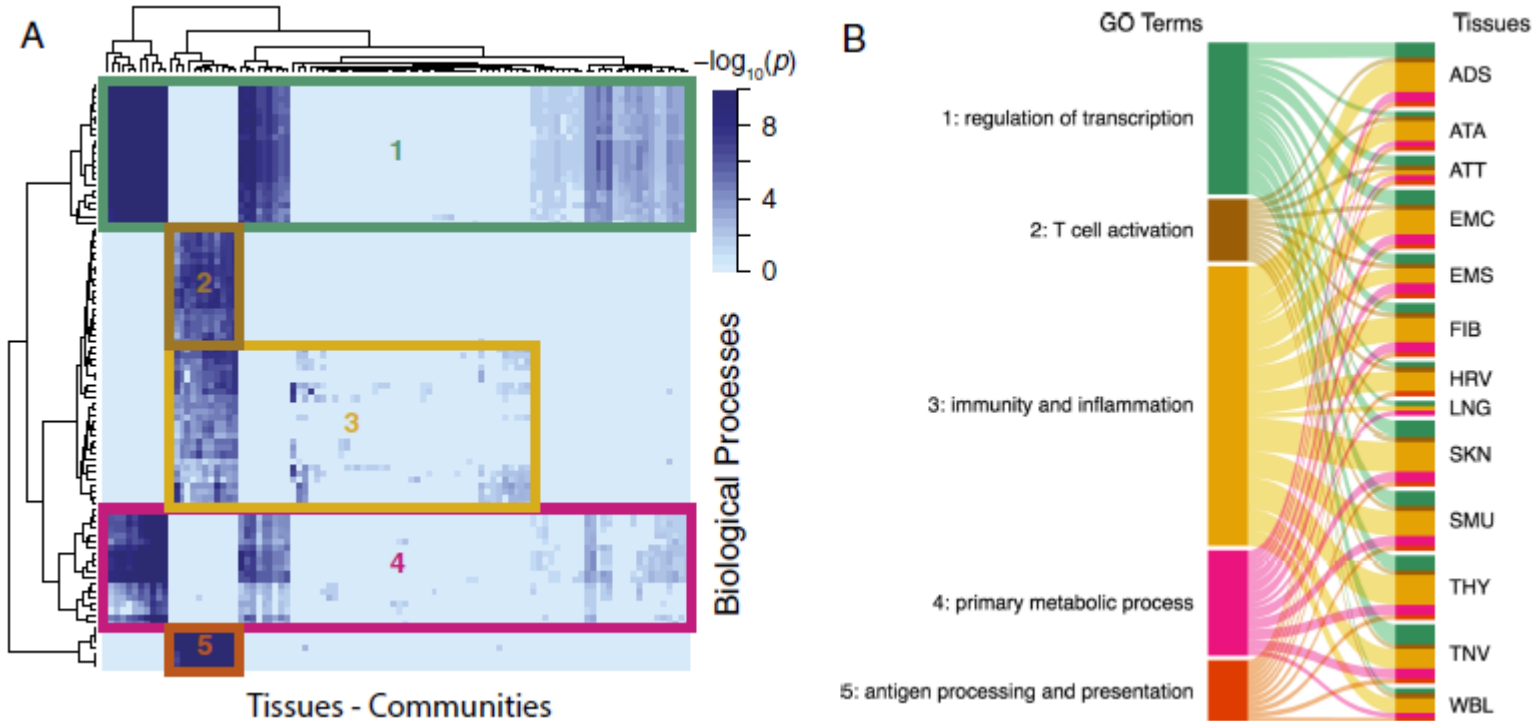
这表明在每个网络中，群体是紧密聚集的，连接来自同一社区的节点的边缘密度远高于链接来自不同社区的节点的边缘密度。



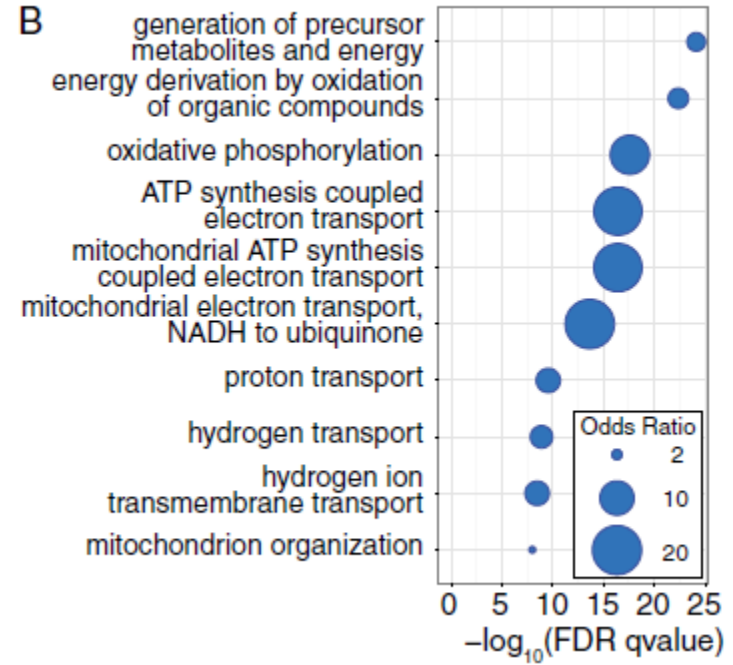
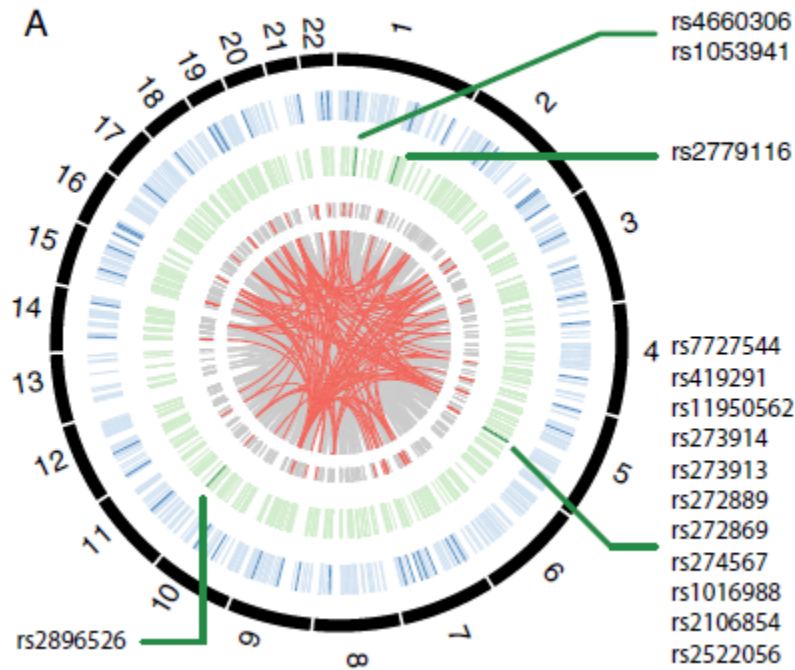
86%的群体中位数绝对相关系数低于0.2，这表明我们的群体不是由基因表达相关驱动的（除了基因很少的群体），因此不可能使用共表达网络方法。这表明多个顺式和反式eQTL SNP对功能相关的基因组的遗传影响，驱动这些群体的组织和结构。



二、eQTL群体结构反映了共享和组织特异性生物学功能



当比较组织之间的群体时，我们发现许多群体具有跨组织功能丰富的共同模式，增强了在这些群体中的SNP的多效性。



A: 心脏左心室群体86的圆环图。黑色代表染色体，蓝色代表基因，绿色代表SNPs，SNP-基因所有的关联是灰色，eQTL与参与细胞呼吸的基因或涉及代谢的SNP的关联是红色

B: 心脏左心室群体86含有大量基因与细胞呼吸有关

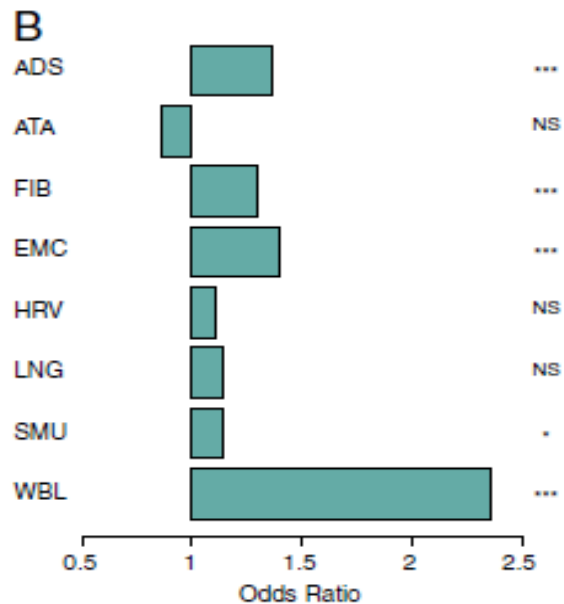
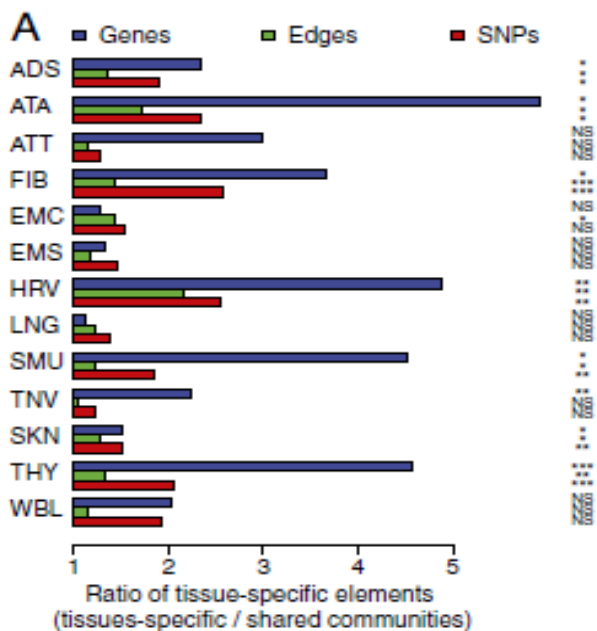


- 根据上图发现**TS**群体包含涉及**TS**功能过程的基因，例如心脏左心室中的细胞呼吸或食道肌肉平滑肌收缩。对于这些群体的组织特异性有一个合理的机制解释：使用来自**the Roadmap Epigenome Project**的八个组织中的数据分析，我们发现这些**TS**群体中的**TS eQTL SNP**富集于该组织独特的活性染色质区域。这表明这些群体的组织是由特定**SNP**周围的染色质区域的表观遗传活化驱动的，并且这些**SNP**以顺式和反式作用对功能相关基因的表达发挥遗传作用，这些基因在各自的组织水平过程中具有重要作用。



三、组织特异性社区的生物学特征

- 鉴于遗传变体存在于所有组织类型中，组织特异性表观遗传活化可能影响其对基因表达的调节作用。因此，又研究他们观察到的组织特异性群体相关的染色质状态变化。



正如预期的那样，我们观察到在所有的组织中，**TS**群体往往比共享群体具有更高比例的**TS**基因。

通过比较**TS**和共享群体，我们发现在八个组织中的五个组织中，**TS**群体的**TS SNP**比共享群体中的**TS SNP**更有可能显示**TS**染色质激活。



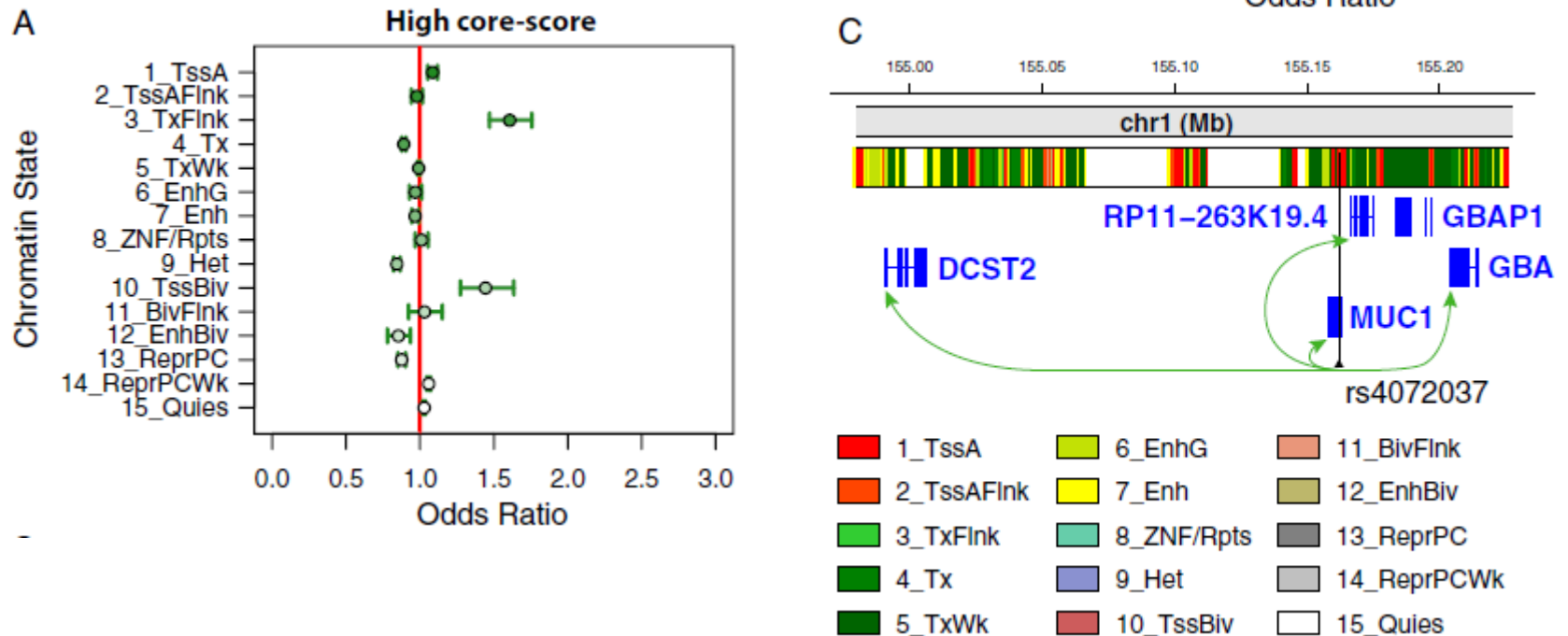
四、整体中心和群体核心反映不同的**SNP**功能

许多研究表明，**eQTL SNP**在调控区域中超表达，但并非所有**eQTL SNP**都具有明显的调节作用。从我们对**eQTL**网络的结构分析中出现的一个自然假说是可能存在与**SNP**的各种调节作用相关联的网络属性。具体来说，人们可能会期望在整体或局部更高度连接的**SNP**将具有与网络外围的**SNP**不同的调控作用。



Community Cores

core-score反映了其群体内每个SNP的局部中心地位。

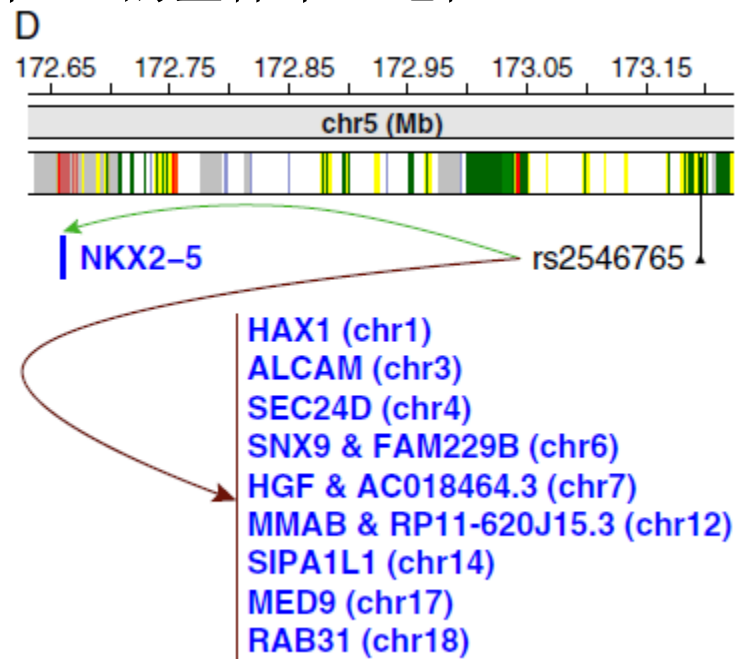
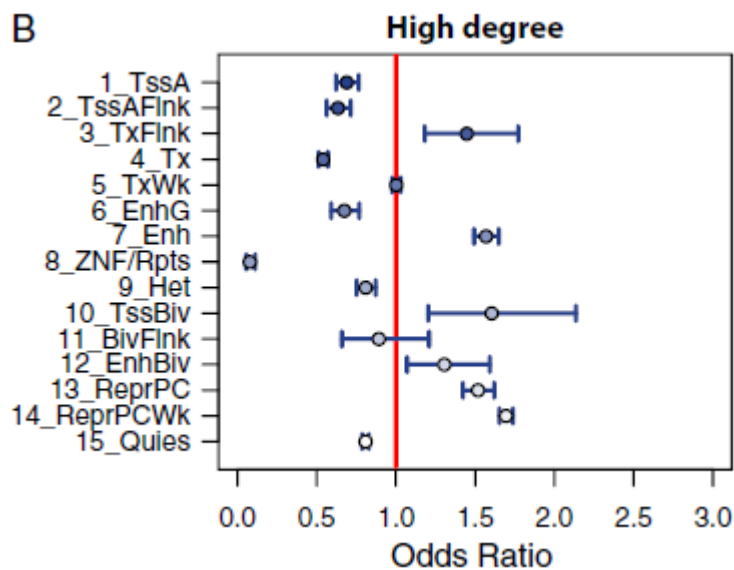


我们发现核心SNP富集启动区，群体中心丰富了接近转录起始位点的活性染色质区域



Global Hubs

SNP degree反映了整个网络中SNP的整体中心地位



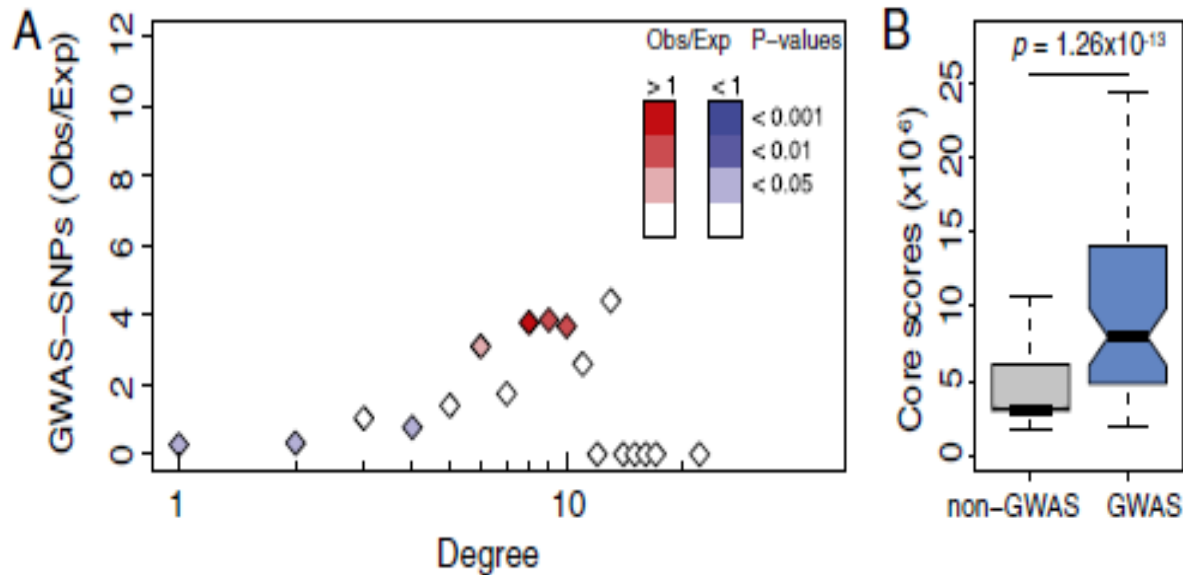
整体中心通常更有可能在非基因增强子和多梳抑制区域内



- 我们发现，整体中心和群体中心与活跃的调控区域都有不同的关联：整体中心优先位于远端调控区域（增强子），而核心**SNP**在近端调控因子（启动子）中比较多。这是eQTL网络结构提供对生物过程的深入了解和跨组织重现的的深入了解的另一个原因。



五、群体中心中的SNP与特征和疾病表型有关。



来自GWAS的性状相关变体的分布在13个组织中也是高度一致的，GWAS SNP富含中级（5-10）网络程度，低度（1-2）较少。群体中心中GWAS SNP的重大超表达表示，进一步说明了这个网络结构的优越。



结论：该网络方法提供了具有不同解释力的框架，其可以作为进一步探索基因型和表型之间的联系的基础。

启发：可以通过eQTL网络研究杂交水稻的性状。

存在的问题：

1. 不能排除在顺式和反eQTL分析中进行的大量统计学检验可能导致一些个体eQTL关联不明显
2. 描述的网络模型不能完全解决弱效应变量 如何确定复杂特征和疾病的问题。