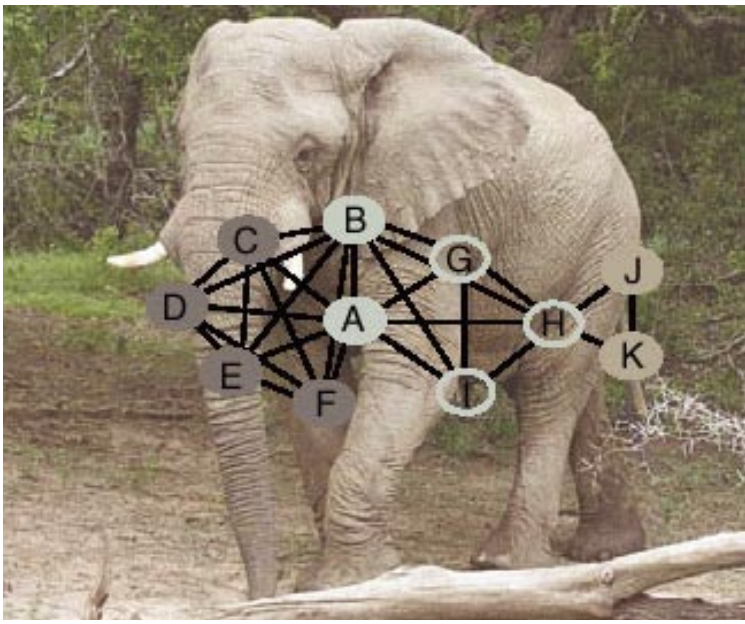
一个将当前生物科学知识全面系统地联系在一起的超级数据库网络

如果盲人摸象这个谚语中的盲人是科学家的话，他们或许就不会闹出那么大的笑话来了。因为随着对大象表观基因组、基因组、转录组、蛋白质组以及糖组越来越精确的分析，大量的数据库将使这些结果的存取变得十分简单，同时也会鼓励越来越多的人们来从事这项研究。但即便如此，我们仍需要知道各部分是怎样协调运作从而组成一个完整的大象的。

目前的难题是怎样将不同来源的知识，信息整合在一起，以了解每一部分在整体中所起的作用。而实现这一目标的瓶颈不再是数据的匮乏，而是缺少灵活的计算方法以及经济支持来将这些数据整合进一个超级数据库。目前许多工作都致力于对基因组结构，基因序列以及表达谱所联系的临床表型的研究。为得到真实可信的信息，功能性网络需要将那些不同的方法整合在一起，例如通过基因组和蛋白质组以及细胞定位相关的数据来丰富相互作用组（interactome）的数据库。

想象一下，随着信息量的扩大，一个真正的超级数据库将出现在我们面前，举例来说，通过这个超级数据库，科学家在进行数据搜索时只需在线输入一个蛋白质名称之后用鼠标点击几下就可以获得海量的相关信息：如该蛋白的基因序列，外部状态，在不同组织的表达谱，在组织中完整的蛋白质组和相互作用组的数据，该蛋白的功能以及与其它蛋白的相互作用方式，转录后的修饰情况，对疾病的影响情况，突变情况以及其它所有可以得到的资源，例如基因敲除小鼠的相关资料等等。

但要实现上述的梦想，我们需要有一个能够将当前数据库连接进一个易于存取的超级网络的方法，以便人们可以最大程度的将个人的知识整合起来，从而解开更多的未知难题。



整体由部分构成

图片来源：*Wildlife Stock Pictures*

原文检索：http://www.nature.com/nmeth/journal/v5/n1/full/nmeth1165.html

葛晶 编译