**全基因组DNA甲基化谱式揭示RA患者CD4+ T细胞基因组异常中国的汉族RA患者CD4+T细胞的**

RA是关节慢性炎症的自身免疫性疾病。最新的证据表明表观遗传学改变参与RA的发病。了解在RA CD4T细胞下调的DNA甲基化的程度和性质，进行了12个RA病人与12个相应的健康人的CD4T细胞的全基因组甲基化比较研究。胞嘧啶的甲基化状态用Illumina甲基化450K芯片进行量化。DNA甲基化分析显示了在RA CD4T细胞的383个高表达和785低表达的甲基化基因(P<3.4×10-7)。基因本体论分析表明，在RA的发病机制中，基因启动子的剪接和蛋白修饰可能在RA的发病机制中起重要作用。此外，结果表明人的HLA区域包括HLA-DRB6, HLA-DQA1 和HLA-E经常过度甲基化，但RA患者HLA-DQB1基因在CpG岛区高甲基化在CpG隐蔽区域低甲基化。在MHC区域之外，HDAC4, NXN, TBCD 和TMEM61是最高甲基化表达的基因，而ITIH3,TCN2, PRDM16, SLC1A5 和 GALNT91是最低甲基化表达的基因。基因组甲基化谱显示RA病人的CD4+ T细胞存在DNA甲基化的改变。

RA是自身免疫性疾病，主要攻击滑膜组织产生关节的慢性炎症。RA的发病机制已经被广泛的研究。最重要的病因学源自单核苷酸多态性（SNP）。在SNPs基础上的全基因组关联研究已经发现了多种与RA有关的SNPs。然而，在我们以前的工作中表现的那样，即使是像甲状腺癌高家族性的疾病，几个重要的单核苷酸多态性可能只有有限的预测能力。如预期，大量的拷贝数的变化与RA的易感性表现出显着的关联。随着从基因研究中积累的系统性自身免疫性类风湿疾病的证据，如SLE，痛风和SSc，常常表现出相同的临床特点和基因危险因素。因此，从理论上讲，这些复杂的疾病可能会具有在表观遗传学水平的一些类似的缺陷。

目前估计RA的遗传率为约20%-50%，这在有无抗环瓜氨酸肽抗体的RA患者中是显著不同的，表明表观遗传学因素在RA的病因学中有着重要的角色。此外，SLE和SjS的CD4+ 或 CD8+ T-cells的DNA全基因组甲基化已经表现出了大量的改变。因此，可以假定DNA甲基化在RA的发病中是有意义的。DNA甲基化是最重要的表观遗传学改变。我们以前的研究，DNA甲基化在基因和microRNA表达调节，癌症的开始和进展中发挥重要的作用，在癌症的诊断和转归中发挥重要的角色。

以往的研究表明几个重要的免疫相关基因在RA基因组中异常的甲基化。然而，全基因组DNA甲基化谱的RA病人还是非常有限的，尤其是在亚洲人口和不能进行的公共区域。从候选基因发现更多的RA相关的DNA甲基化区域和识别缺失的RA遗传是一个很大的问题。在现段的研究中，我们进行了12例RA 和相匹配的12名健康人的CD4 + T细胞的全基因组DNA甲基化谱分析。

**患者和对照组**：12例RA病人和12例健康对照进行研究，平均年龄病人是42.83，正常人是43.75 无统计学差异 (p = 0.95)。所有患者均符合美国RA风湿病学会分类标准的。

**分离PBMCs和CD4+ T细胞：**外周血单个核细胞从RA和正常样本血液标本采用密度梯度离心制备（Amersham Biosciences）使用标准的步骤和快速处理培养细胞。CD4 + T细胞从新鲜分离的外周血单个核细胞中通过消耗表达CD8、CD14、CD16、CD56、CD19、CD36，CD123，γ/δT受体的细胞制备，与血型糖蛋白A采用无接触的T细胞分离试剂盒（miltenyibiotec）。CD4+ T细胞纯度是95–98%，用特异性抗体流式细胞仪测定。全基因组DNA用QIAGEN试剂盒制备(Qiagen, Germantown, MD)和使用EZ DNA甲基化酶试剂盒进行亚硫酸氢钠处理(Zymo, Orange, CA).

**人的甲基化450芯片：**

亚硫酸氢钠转换的病人和正常组DNA样本和使用Nanodrop扫描分光光度计量化进行准备和量化。对于每个样品，500 ng的全基因组亚硫酸氢钠转化DNA被变性，粉碎，放大，甲基化信号用humanmethylation450k芯片(Illumina, San Diego, CA, USA)进行检测。

标准的DNA甲基化450K分析渠道（SMAP）来进行甲基化基因芯片分析。基因组工作室（Illumina）是用来产生信号强度与监测内部控制规范（ICN）的P值与和背景减法（BS）。质量控制和标准化用R包的“lumi”进行。位于ChrX和ChrY单核苷酸多态性的探针移出来进行进一步分析。此外，有检测P值> 0.01超过5%的样品也被过滤掉，而其他探头小于5%的样品标记为缺失值的探针（NA）避免以下统计和生物信息学进一步分析的歧义。然后整体信号强度，m值的分布和数量检测网站被用来衡量芯片的质量。显著的异常样品/芯片之前的差异甲基化位点识别删除。信号强度与包装的“lumi”进行颜色偏差的调整和分位数归一化（QN）。最后，β混合分量的归一化（BMIQ）的β值进行了调整，去调整不同类型的探针造成的偏差（I型和II型）。

**数据分析：**

应用主成分分析法和聚类分析法对样本间的相关性进行分析。在主成分分析中，有2个样本被筛选出来，因为它们在主成分分析中明显不同。通过配对t检验的基础上的β值的归一化数据的差异甲基化位点进行鉴定。原始值的错误发现率（FDR）= 0.05，调整为多个测试校正。临床特征与鉴别甲基化位点之间的关联性进行线性回归分析阈值为0.005。用DAVID生物信息学资源进行基因本体论分析。通过字符串推断差异甲基化基因的相互作用。人类的grch37/ hg19标记用做生物信息学分析和作为结果表述。所有的方法和分析在R进行。数据存放在基因表达数据库(GEO accession: GSE71841)。

**结果：**

我们研究RA患者和正常人的全基因组CD4+Tcell的DNA甲基化，采用基于微珠芯片的高通量筛选法，允许482421个CpG位点跨越21231个基因的启动子区域（99% RefSeq基因）同时筛选。24个生物复制样本（12个RA和12个匹配的健康对照组），收集和登记的项目。（表1）。检测值0.01以上进行过滤删除1067个探针数据集。在至少5%的样品中用一颗珠子数< 3把443探针从数据集上取下。29021探针SNPs（dbSNP数据库版本：142）被移除降低分析的偏差。11245和8510探针被删除，因为多序列比对或位于X-染色体或染色体Y，最终，435226个探针被保存在24个样本。

为了确保我们的研究细胞均来源于CD4+T细胞，我们预计我们的甲基化信号进入PC1和PC2的全血细胞尺寸。分析表明，我们的样品明显聚集在CD4+细胞和疏远其他种类的细胞，如CD8 + T细胞，CD14 +，CD19 +等。表明样品采集过程中成功地制备出的样品（图1A）。更重要的是，确保发现的不同的甲基化模式在我们的研究中不受RA患者和对照组的T细胞亚群的潜在差异的影响，我们发现许多基因与已知的特定的T细胞亚群的相关的甲基化谱，如IL4、IL13（Th2），IFNγ（Th1）和IL17F（Th17）。在病例和对照之间这些CpG位点没有发现差异显著，提示RA患者和对照组在T细胞亚群之间无差异。

同时，PCA分析我们的甲基化450K，结果表明PC1和PC2解释总方差的29.9%和14%， 13大主成分分别可以解释80%的方差。这些结果表明，有限的临床或人口信号和数据的信息对为进一步的生物信息学和生物统计学的分析是可信的（补充图1）。

**RA的全基因组DNA甲基化谱**

我们确定了在RA与正常对照组相比CD4+ T细胞的810个低甲基化，392个过甲基化CpG位点，代表了在RA患者低甲基化和过甲基化的785个和383个基因P<3.4×10-7。（配对t检验，率＜0.05，补充）。基于显著差异甲基化位点的聚类分析显示RA和正常对照组之间的明显分离（图1B）。在RA中更多的低甲基化CpG位点超过过甲基化位点表明全基因组的低甲基化。

差异甲基化基因的相互作用是基于字符串10.0，结果表明DMGs是高度的相互关联的而不是功能分离（图2）。基因本体论分析表明，在RA患者选择性剪接（P = 1.2×10-7，FDR）、磷酸化蛋白（P = 1.7×10-2，FDR）明显异常（表2），提示在RA的发病过程中，由甲基化修饰的剪接和蛋白修饰异常可能在RA的发病中起重要作用。此外，免疫反应（p值= 3.2×10-5）和白细胞单核细胞（p值= 0.02）相关的基因本体论也明显丰富。

更重要的是，结果表明，在RA患者人类白细胞抗原（HLA）区域频繁的低甲基化，包括HLA-DRB6 (P=6.61×10-10), HLA-DQA1 (P=7.09×10-9)、HLA-E (P=3.24×10-7)，然而，HLA-DQB1在CpG岛区的过甲基化和CpG平台区的低甲基化表现出不同的甲基化表象。MHC区域的外面，在RA最过甲基化的基因包括HDAC4 (P=1.47×10-7), NXN (P=5.5×10-9), TBCD (P=4.48×10-8) 和TMEM61 (P=1.7×10-7)，而最显着的低甲基化的基因包括ITIH3 (P=1.16×10-7), TCN2 (P=1.57×10-8), PRDM16 (P=3.1×10-9), SLC1A5 (P=2.94×10-7) and GALNT9 (P=8.26×10-9)。

**DNA甲基化和疾病特点的关系：**

大量的病人临床特点记录下来。确定临床相关的甲基化位点，将对RA的病理机制提供重要的见解，并有临床应用价值。临床特点和甲基化位点的关联性进行分析，我们发现甲基化水平OR5A2 (cg02981094, P=2.6×10-4), ALDH9A1 (cg03984859, P=2.8×10-4) and C5orf32 (cg02070114, P=2.2×10-4)是与RA的疾病进程相关的。

除外，ZC3H11A(cg02337583)的甲基化水平与RA患者类风湿因子相关 (P=8.9×10-4)。除外，OAS2 (cg00085448)的甲基化水平与RA病人的的HZPG密切联系(P=4.1×10-4)。C16orf71 (cg04705084), LOC100129716 (cg00598143) and miR-762 (cg02558026)与DAS 28密切相关P-value of 5.8×10-3, P-value of 5.8×10-3。五个位点包括SLC38A8 (cg01740650, P=3.0×10-3), C18orf19 (cg00448482, P=3.0×10-3), COL18A1 (cg04760448, P=1.9×10-3), BAT3 (cg05649229, P=4.9×10-3) ，PLD3 (cg07071106, P=4.4×10-3) 与ESR显著相关。最后，我们发现HSPA12A (cg06942850)与压痛关节数密切相关P-值3.2×10-3.。

**讨论：**

总之，我们应用汉族RA患者CD4+T细胞全基因组DNA甲基化变化的Illumina甲基化450K芯片。基于严格的测量和分析，1202个CpG位点的DNA甲基化水平在RA和正常组的CD4+T细胞存在显著差异。基因ontology分析及基因互作分析表明，这些基因并非相互独立的个体而是功能相关，相互作用，可能共同在RA的发病中发挥功能。全基因组甲基化数据显示在RA患者中具有更多的低甲基化位点，与以往的文献报道一致。

与通常含有千个不同甲基化位点的肿瘤基因甲基化改变相比，系统性自身免疫性风湿性疾病似乎只有几个不同的甲基化区域。我们的甲基化数据主成分分析也显示在RA和正常组没有明显的分离，在RA和正常人之间CD4+T细胞不会有很多不同的甲基化区域。Kazuhisa及其同事用人类识别1,859位点的甲基化450K芯片进行了一项成纤维样滑膜细胞（FLS）和RA全基因组DNA甲基化检测比较差异甲基化位点。在SLE和正常人之间的CD4+发现341差异甲基化位点，在SJS和健康对照CD4 +之间发现753差异甲基化位点。

尽管多种测试修订已经进行。我们还认为仍有大量的不同的甲基化位点的阳性是错误。因此，虽然已经进行了多个测试校正这些文件，我们认为，仍有大量的差异甲基化位点会出现假阳性。因此，RA的发病机制候选差异甲基化位点将是有限的。与Kazuhisa的研究相比，有81个共同的差异甲基化CpG位点可能在RA的发病机制中非常重要。不同种族人群的全基因组甲基化谱可能会为RA患者的甲基化变化提供特定的民族信息。

我们的研究没有在另一个独立的队列执行验证，因为先前的研究表明甲基化450K精确度很高。我们也没有进行基因表达分析，因为甲基化功能的对于复杂疾病的影响，不仅在基因表达，但也在其他一些重要的功能，如替代表达，基因组稳定性和相互作用与遗传变异，如单核苷酸多态性。

另一方面，我们要强调的是RA的表观遗传机制应不仅外周血单个核细胞也对血细胞的组成给予更多的精度。在下一步，我们将在RA完成CD8+、CD17+T细胞全基因组甲基化谱，为每个免疫细胞提供表观遗传的贡献。此外，我们的数据提供了一个在系统性红斑狼疮，痛风和其他自身免疫性疾病比较CD4+甲基化谱的机会。我们发现[Jeffries](http://www.ncbi.nlm.nih.gov/pubmed/?term=Jeffries%20MA%5Bauth%5D)RA和正常高加索人群之间确定了761个差异甲基化CpG位点。

我们比较了我们的和Jeffries的研究中的异常的不同甲基化位点，发现GALNT9在两个研究中都存在，GALNT9在RA病人中过甲基化表达。尽管只有一个基因在两个研究中相同，我们认为这是可以接受的，因为我们和[Jeffries](http://www.ncbi.nlm.nih.gov/pubmed/?term=Jeffries%20MA%5Bauth%5D)’s的研究发现的差异甲基化位点是小样本量和低统计效率的总差异甲基化位点的一小部分。更为显著的是，在我们的研究中，有大量的遗传、环境风险和临床特征的影响，这可能会带来一些不同的研究结果。我们的研究表明，表观遗传学为基础的关联研究或生物标志物的鉴定在不同的人群中确实需要大样本，最终可以找到共享的表观遗传生物标志物。

**结论：**

全基因组DNA甲基化模式显示RA患者CD4 + T细胞DNA甲基化状态存在显著改变。