

# Genetic and epigenetic fine mapping of causal autoimmune disease variants

*Mikhail Dozmorov*

*November 1, 2014*

Genetic and epigenetic fine mapping of causal autoimmune disease variants  
(<http://www.nature.com/nature/journal/vaop/ncurrent/pdf/nature13835.pdf>)

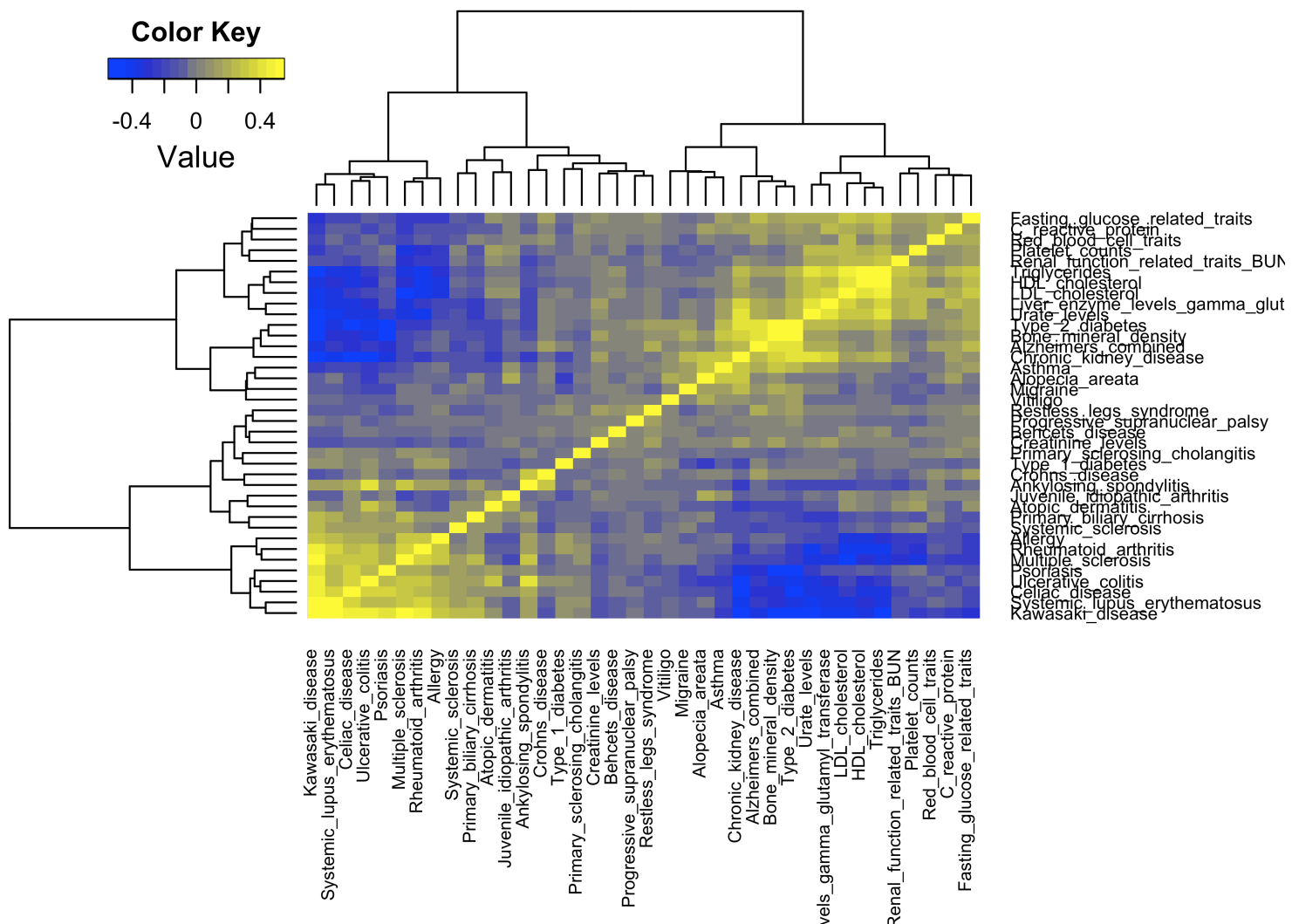
Supplemental table 1 (<http://www.nature.com/nature/journal/vaop/ncurrent/extref/nature13835-s1.xls>) has genomic coordinates of disease-associated SNPs.

## Analysis of TFBSs

Out of all regulatory datasets, we select only TFBSs.

```
## [1] 1954 39
```

```
## [1] 1259 38
```



**Text mining question 1:** Are the diseases within a cluster share stronger literature similarity than the diseases between the clusters? To answer, we need literature similarity scores for each pair, then split the pairs into cluster-specific groups and compare score distributions with what can be expected by chance, calculating the p-values for it. *Expected answer:* Diseases within each cluster are related to each other by literature findings stronger than could be expected by chance. Diseases between the clusters are not related to each other by literature findings, and this also may be statistically significant.

The top 10 pairs of disease-associated SNPs are most similar with each other.

```

##
## -----
-----
##          Disease 1                      Disease 2          Corr. coeffic
ient
## -----
-----
##          HDL_cholesterol                Triglycerides        0.5484

##
##          Kawasaki_disease                Systemic_lupus_erythematosus    0.5352

##
##          Bone_mineral_density            Type_2_diabetes        0.5268

##
##          Kawasaki_disease                Multiple_sclerosis        0.501

##
##          Kawasaki_disease                Rheumatoid_arthritis      0.4775

##
##          Celiac_disease                  Kawasaki_disease         0.4754

##
##          LDL_cholesterol                 Triglycerides            0.4743

##
##          Kawasaki_disease                Ulcerative_colitis       0.4661

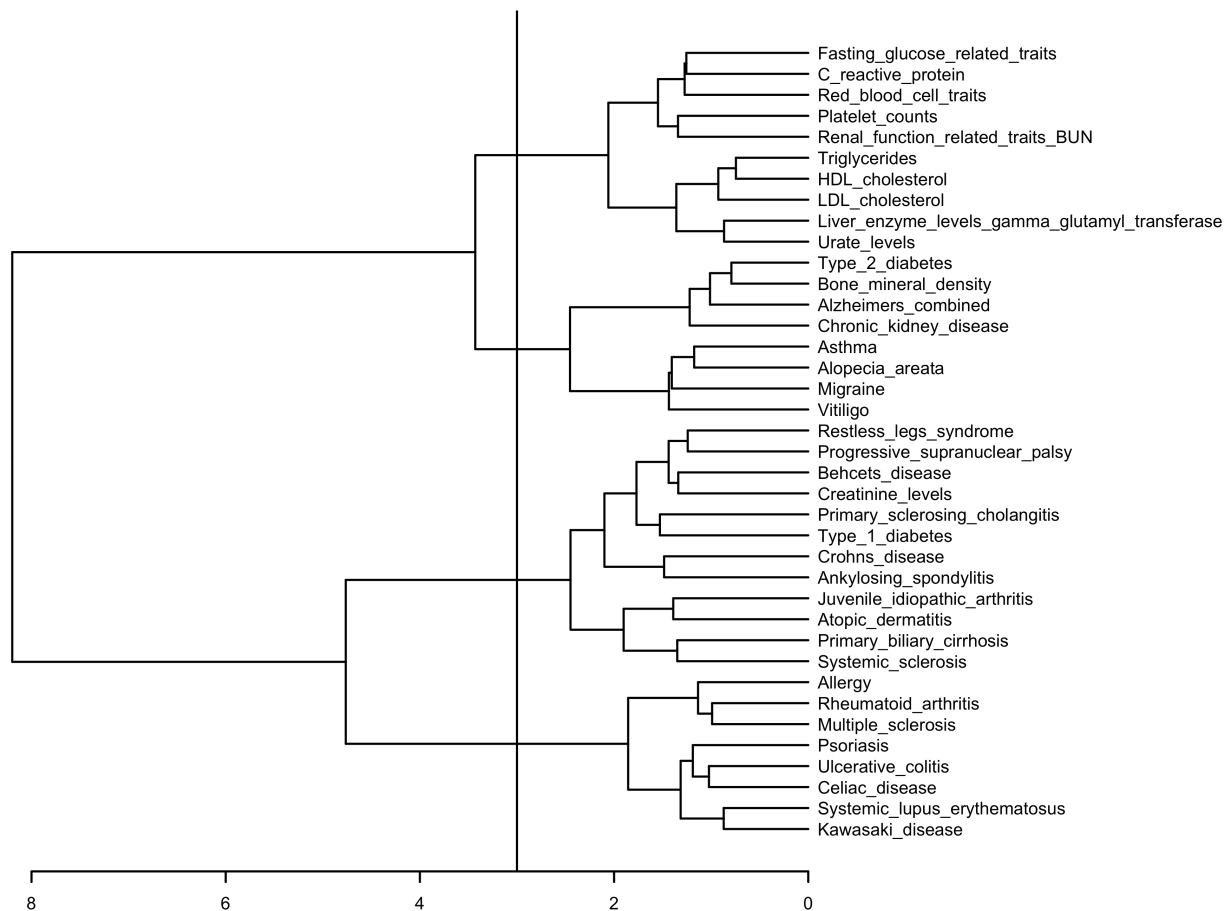
##
## Liver_enzyme_levels_gamma_glutamyl_transferase    Urate_levels            0.4191

##
##          Alzheimers_combined              Bone_mineral_density     0.4149

## -----
-----

```

The similarity dendrogram can be divided into separate groups:



```
## Cluster01 has 8 members
## Kawasaki_disease
## Systemic_lupus_erythematosus
## Celiac_disease
## Ulcerative_colitis
## Psoriasis
## Multiple_sclerosis
## Rheumatoid_arthritis
## Allergy
##
## Cluster02 has 12 members
## Systemic_sclerosis
## Primary_biliary_cirrhosis
## Atopic_dermatitis
## Juvenile_idiopathic_arthritis
## Ankylosing_spondylitis
## Crohns_disease
## Type_1_diabetes
## Primary_sclerosing_cholangitis
## Creatinine_levels
## Behcets_disease
## Progressive_supranuclear_palsy
## Restless_legs_syndrome
##
## Cluster03 has 8 members
## Vitiligo
## Migraine
## Alopecia_areata
## Asthma
## Chronic_kidney_disease
## Alzheimers_combined
## Bone_mineral_density
## Type_2_diabetes
##
## Cluster04 has 10 members
## Urate_levels
## Liver_enzyme_levels_gamma_glutamyl_transferase
## LDL_cholesterol
## HDL_cholesterol
## Triglycerides
## Renal_function_related_traits_BUN
## Platelet_counts
## Red_blood_cell_traits
## C_reactive_protein
## Fasting_glucose_related_traits
##
```

The “Enrichment 1/2” columns show the average p-values of the group-specific SNPs-regulatory associations. A “-” sign indicates that an association is underrepresented. The “p-value” column shows whether the difference in the associations between the groups is statistically significantly different.

[illegible]

```
##
##
##      wgEncodeHaibTfbsGm12892Pol2V0416102PkRep1      GM12892 Pol2 v041610.2      0.0001446
0.7113  0.004147
##
##      ChIP-seq Peaks Rep 1 from
##
##      ENCODE/HAIB
##
##      wgEncodeSydhTfbsGm12878Chd1a301218aIggmusPk      GM12878 CHD1 IgG-mus ChIP-seq 0.0001594
0.7835  0.004317
##
##      Peaks from ENCODE/SYDH
##
##      wgEncodeHaibTfbsGm12878Pol24h8Pcr1xPkRep1      GM12878 Pol2-4H8 PCR1x      0.0002028
0.965   0.05724
##
##      ChIP-seq Peaks Rep 1 from
##
##      ENCODE/HAIB
##
##      wgEncodeHaibTfbsGm12878Pmlsc71910V0422111PkRep2      GM12878 PML v042211.1      0.0001479
0.7001   8.3e-05
##
##      ChIP-seq Peaks Rep 2 from
##
##      ENCODE/HAIB
##
## -----
##
## [1] "c1 vs. c3 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 110"
##
## -----
##
##      Row.names      V2      c1
c3      adj.P.Val
## -----
##
##      wgEncodeHaibTfbsGm12878Mta3sc81325V0422111PkRep2      GM12878 MTA3 v042211.1      2.011e-05
-0.03223  9.655e-06
##
##      ChIP-seq Peaks Rep 2 from
##
##      ENCODE/HAIB
##
##
##      wgEncodeSydhTfbsGm12878Pol2IggmusPk      GM12878 Pol2 IgG-mus ChIP-seq 9.774e-06
```

```

-0.06673  0.0006752
##
## Peaks from ENCODE/SYDH

##
## wgEncodeSydhTfbsGm18505NfkbTnfaIggrabPk GM18505 NFKB IgG-rab TNFa 3.594e-06
-0.4501  0.03447
##
## ChIP-seq Peaks from

##
## ENCODE/SYDH

##
## wgEncodeSydhTfbsGm12891Pol2IggmusPk GM12891 Pol2 IgG-mus ChIP-seq 6.286e-05
-0.1467  0.01099
##
## Peaks from ENCODE/SYDH

##
## wgEncodeHaibTfbsGm12878Nficsc81335V0422111PkRep1 GM12878 NFIC v042211.1 6.767e-05
-0.1369  8.824e-06
##
## ChIP-seq Peaks Rep 1 from

##
## ENCODE/HAIB

##
## wgEncodeHaibTfbsGm12892Pol2V0416102PkRep1 GM12892 Pol2 v041610.2 0.0001446
-0.09939 0.0005075
##
## ChIP-seq Peaks Rep 1 from

##
## ENCODE/HAIB

##
## wgEncodeHaibTfbsGm12878Nfatc1sc17834V0422111PkRep2 GM12878 NFATC1 v042211.1 9.584e-05
-0.1597  8.211e-05
##
## ChIP-seq Peaks Rep 2 from

##
## ENCODE/HAIB

##
## wgEncodeHaibTfbsGm12878Pmlsc71910V0422111PkRep2 GM12878 PML v042211.1 0.0001479
-0.1222  8.824e-06
##
## ChIP-seq Peaks Rep 2 from

##
## ENCODE/HAIB

##
## wgEncodeSydhTfbsGm18505Pol2IggmusPk GM18505 Pol2 IgG-mus ChIP-seq 0.0002114
-0.1043  0.0003008
##
## Peaks from ENCODE/SYDH

```



```

##
##      wgEncodeHaibTfbsGm12892Pol124h8V0416102PkRep1      GM12892 Pol2-4H8 v041610.2      0.0003613
##      -0.06464      0.000142
##
##                                     ChIP-seq Peaks Rep 1 from
##
##                                     ENCODE/HAIB
## -----
##
## [1] "c1 vs. c4 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 103"
##
## -----
##
##      Row.names      V2      c1
##      c4      adj.P.Val
## -----
##
##      wgEncodeSydhTfbsGm18505NfkbTnfaIggrabPk      GM18505 NFKB IgG-rab TNFa      3.594e-06
##      -0.775      0.03454
##
##                                     ChIP-seq Peaks from
##
##                                     ENCODE/SYDH
##
##
##      wgEncodeSydhTfbsGm12878Pol2IggmusPk      GM12878 Pol2 IgG-mus ChIP-seq 9.774e-06
##      -0.3918      0.001616
##
##                                     Peaks from ENCODE/SYDH
##
##
##      wgEncodeHaibTfbsGm12878Mta3sc81325V0422111PkRep2      GM12878 MTA3 v042211.1      2.011e-05
##      -0.2356      3.528e-05
##
##                                     ChIP-seq Peaks Rep 2 from
##
##                                     ENCODE/HAIB
##
##
##      wgEncodeHaibTfbsGm12878Nficsc81335V0422111PkRep1      GM12878 NFIC v042211.1      6.767e-05
##      -0.1347      5.071e-06
##
##                                     ChIP-seq Peaks Rep 1 from
##
##                                     ENCODE/HAIB
##
##
##      wgEncodeSydhTfbsGm12891Pol2IggmusPk      GM12891 Pol2 IgG-mus ChIP-seq 6.286e-05
##      -0.5071      0.01766

```

```
## Peaks from ENCODE/SYDH

##
## wgEncodeHaibTfbsGm12878Pmlsc71910V042211PkRep2 GM12878 PML v042211.1 0.0001479
-0.3198 9.171e-06
## ChIP-seq Peaks Rep 2 from

## ENCODE/HAIB

##
## wgEncodeHaibTfbsGm12878Nfatc1sc17834V042211PkRep2 GM12878 NFATC1 v042211.1 9.584e-05
-0.4966 0.0001887
## ChIP-seq Peaks Rep 2 from

## ENCODE/HAIB

##
## wgEncodeHaibTfbsGm12878Mta3sc81325V042211PkRep1 GM12878 MTA3 v042211.1 0.000339
-0.2 2.35e-05
## ChIP-seq Peaks Rep 1 from

## ENCODE/HAIB

##
## wgEncodeHaibTfbsGm12878Pol24h8Pcr1xPkRep2 GM12878 Pol2-4H8 PCR1x 0.0004474
-0.1903 0.004696
## ChIP-seq Peaks Rep 2 from

## ENCODE/HAIB

##
## wgEncodeHaibTfbsGm12892Pol2V0416102PkRep1 GM12892 Pol2 v041610.2 0.0001446
-0.6038 0.00179
## ChIP-seq Peaks Rep 1 from

## ENCODE/HAIB

## -----
##
## [1] "c2 vs. c3 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 0"
##
## -----
##
## Row.names V2 c2
c3 adj.P.Val
```

```

-----
## wgEncodeHaibTfbsMcf7Hdac2sc6296V0422111PkRep2 MCF-7 HDAC2 v042211.1 1.024e-08 -
0.9111 0.6711
## ChIP-seq Peaks Rep 2 from

## ENCODE/HAIB

##
## wgEncodeHaibTfbsA549Tead4sc101184V0422111PkRep1 A549 TEAD4 v042211.1 ChIP-seq 2.584e-07
0.7003 0.6885
## Peaks Rep 1 from ENCODE/HAIB

##
## wgEncodeSydhTfbsMcf7Gata3sc269UcdPk MCF-7 GATA3 SC269 UC Davis 5.462e-06 -
0.9907 0.6676
## ChIP-seq Peaks from

## ENCODE/SYDH

##
## wgEncodeSydhTfbsH1hescSuz12UcdPk H1-hESC SUZ12 UC Davis 5.736e-06
0.3309 0.7192
## ChIP-seq Peaks from

## ENCODE/SYDH

##
## wgEncodeHaibTfbsMcf7MaxV0422111PkRep2 MCF-7 Max v042211.1 ChIP-seq 0.0001833
0.8041 0.6885
## Peaks Rep 2 from ENCODE/HAIB

##
## wgEncodeHaibTfbsMcf7Tcf12V0422111PkRep1 MCF-7 TCF12 v042211.1 0.002079
0.9427 0.6192
## ChIP-seq Peaks Rep 1 from

## ENCODE/HAIB

##
## wgEncodeSydhTfbsGm12878JundIggrabPk GM12878 JunD IgG-rab ChIP-seq 0.02161 -
0.9099 0.7584
## Peaks from ENCODE/SYDH

##
## wgEncodeHaibTfbsGm12878Mta3sc81325V0422111PkRep2 GM12878 MTA3 v042211.1 0.7491 -
0.03223 0.4004
## ChIP-seq Peaks Rep 2 from

```

```
##                                     ENCODE/HAIB

##
##      wgEncodeHaibTfbsHl60Pol24h8V0422111PkRep2          HL-60 Pol2-4H8 v042211.1      0.0397
-0.657      0.4988
##                                     ChIP-seq Peaks Rep 2 from

##                                     ENCODE/HAIB

##
##      wgEncodeHaibTfbsGm12892Pol24h8V0416102PkRep1        GM12892 Pol2-4H8 v041610.2    0.4503   -
0.06464     0.3761
##                                     ChIP-seq Peaks Rep 1 from

##                                     ENCODE/HAIB

## -----
-----
##
## [1] "c2 vs. c4 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 0"
##
## -----
-----
##                               Row.names                                V2                                c2
c4      adj.P.Val
## -----
-----
##      wgEncodeHaibTfbsMcf7Hdac2sc6296V0422111PkRep2        MCF-7 HDAC2 v042211.1      1.024e-08 0
.7626      0.8601
##                                     ChIP-seq Peaks Rep 2 from

##                                     ENCODE/HAIB

##
##      wgEncodeHaibTfbsA549Tead4sc101184V0422111PkRep1      A549 TEAD4 v042211.1 ChIP-seq 2.584e-07 0
.8079      0.8601
##                                     Peaks Rep 1 from ENCODE/HAIB

##
##      wgEncodeSydhTfbsMcf7Gata3sc269UcdPk                  MCF-7 GATA3 SC269 UC Davis  5.462e-06 0
.9893      0.8463
##                                     ChIP-seq Peaks from

##                                     ENCODE/SYDH

##
```

##	wgEncodeSydhTfbsH1hesCSuz12UcdPk	H1-hESC SUZ12 UC Davis	5.736e-06	0
.8414	0.8601			
##		ChIP-seq Peaks from		
##		ENCODE/SYDH		
##				
##	wgEncodeHaibTfbsMcf7MaxV0422111PkRep2	MCF-7 Max v042211.1 ChIP-seq	0.0001833	0
.8538	0.8601			
##		Peaks Rep 2 from ENCODE/HAIB		
##				
##	wgEncodeHaibTfbsMcf7Tcf12V0422111PkRep1	MCF-7 TCF12 v042211.1	0.002079	0
.8988	0.8204			
##		ChIP-seq Peaks Rep 1 from		
##		ENCODE/HAIB		
##				
##	wgEncodeSydhTfbsGm12878JundIggrabPk	GM12878 JunD IgG-rab ChIP-seq	0.02161	-
0.9997	0.9213			
##		Peaks from ENCODE/SYDH		
##				
##	wgEncodeHaibTfbsH160Pol124h8V0422111PkRep2	HL-60 Pol12-4H8 v042211.1	0.0397	-
0.7606	0.7664			
##		ChIP-seq Peaks Rep 2 from		
##		ENCODE/HAIB		
##				
##	wgEncodeHaibTfbsGm12878Pol124h8Pcr1xPkRep2	GM12878 Pol12-4H8 PCR1x	0.3719	-
0.1903	0.8628			
##		ChIP-seq Peaks Rep 2 from		
##		ENCODE/HAIB		
##				
##	wgEncodeHaibTfbsGm12878Nficsc81335V0422111PkRep1	GM12878 NFIC v042211.1	0.5832	-
0.1347	0.6952			
##		ChIP-seq Peaks Rep 1 from		
##		ENCODE/HAIB		
##	-----			
##	-----			

## [1] "c3 vs. c4 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 0"

##

## -----

## Row.names V2 c3

c4 adj.P.Val

## -----

-----

## wgEncodeSydhTfbsK562Pol2s2StdPk K562 Pol2 S2 Standard 0.143 -0  
.8604 0.2013

## ChIP-seq Peaks from

## ENCODE/SYDH

##

## wgEncodeSydhTfbsImr90Mazab85725IggrabPk IMR90 MAZ (ab85725) IgG-rab 0.1393 0.  
6184 0.6938

## ChIP-seq Peaks from

## ENCODE/SYDH

##

## wgEncodeAwgTfbsSydhH1hesccSin3anb6001263IggrabUniPk H1-hESC TFBS Uniform Peaks of 0.2549 0.  
9992 0.221

## SIN3A\_(NB600-1263) from

## ENCODE/Stanford/Analysis

##

## wgEncodeUwTfbsBjCtcfStdHotspotsRep1 BJ CTCF TFBS ChIP-seq 0.2954 -0  
.8671 0.1851

## Hotspots 1 from ENCODE/UW

##

## wgEncodeUwTfbsHmfCtcfStdHotspotsRep1 HMF CTCF TFBS ChIP-seq 0.2174 0.  
8316 0.1651

## Hotspots 1 from ENCODE/UW

##

## wgEncodeHaibTfbsHct116Rad21V0422111PkRep1 HCT-116 RAD21 v042211.1 0.334 -0  
.9289 0.1625

## ChIP-seq Peaks Rep 1 from

## ENCODE/HAIB

##

## wgEncodeHaibTfbsEcc1Rad21V0422111PkRep2 ECC-1 RAD21 v042211.1 0.3424 -0

```

.9585    0.1651
##
ChIP-seq Peaks Rep 2 from

##
ENCODE/HAIB

##
##
wgEncodeUwTfbsHbmecCtcfStdHotspotsRep1    HBMEC CTCF TFBS ChIP-seq    0.32    0.
9727    0.1651
##
Hotspots 1 from ENCODE/UW

##
##
wgEncodeHaibTfbsA549Ctcfsc5916Pcr1xEtoh02PkRep1    A549 CTCF 5916 EtOH PCR1x    0.3365 -0.
.9842    0.1625
##
ChIP-seq Peaks Rep 1 from

##
ENCODE/HAIB

##
##
wgEncodeAwgTfbsSydhH1hescSuz12UcdUniPk    H1-hESC TFBS Uniform Peaks of 0.3232 0.
9742    0.4938
##
SUZ12 from

##
ENCODE/USC/Analysis

## -----
-----

```

**Text mining question 2:** Are the terms associated stronger with the diseases in one vs. the other cluster based on the literature strength? Are the terms themselves related based on the literature? *Expected answer:* Yes, the literature associations should confirm the relationships.

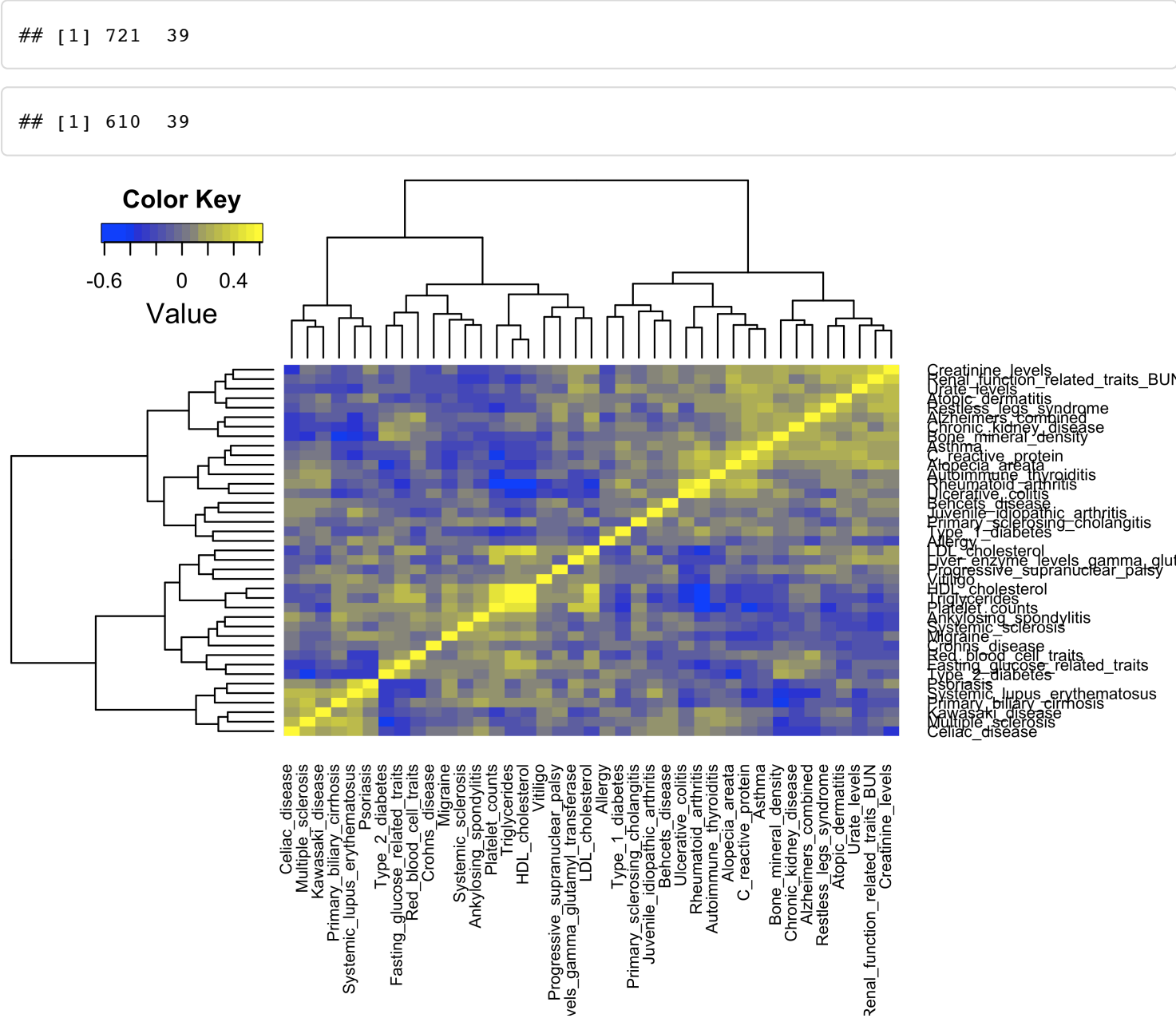
# Summary

1. There are 4 clusters. The first cluster drives all the differences.

C1	C2	C3	C4
C1	Cell types: Gm12878 Reg: NFkB, Pol2, MTA3, NFIC, NFATC1	Cell types: Gm12878 Reg: NFkB, Pol2, MTA3, NFIC, NFATC1	Cell types: Gm12878 Reg: NFkB, Pol2, MTA3, NFIC, NFATC1
C2		Nothing significant	Nothing significant
C3			Nothing significant
C4			

# Analysis of histone marks

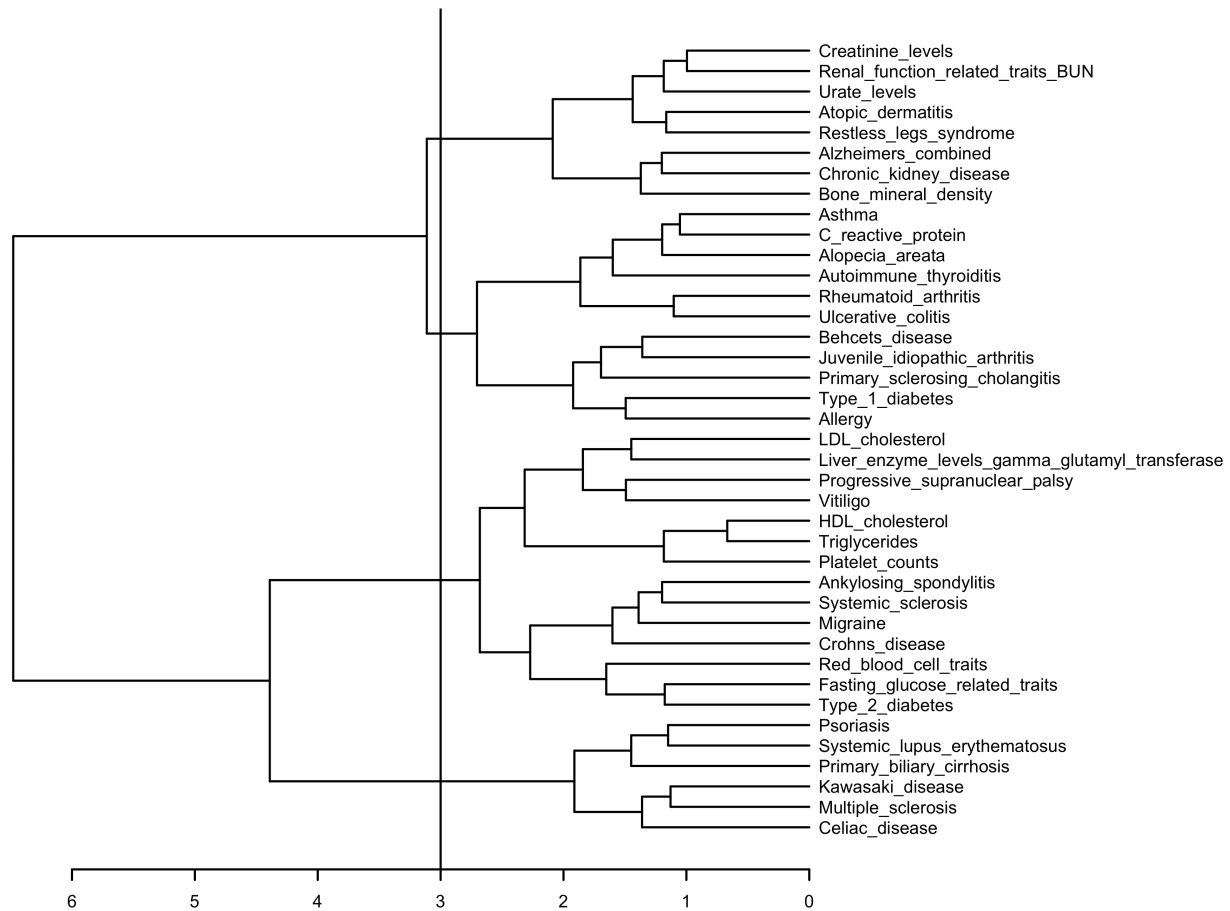
Out of all regulatory datasets, we select only histone marks





##			
##	-----		
##	Disease 1	Disease 2	Corr. coefficient
##	-----		
##	HDL_cholesterol	Triglycerides	0.621
##			
##	Rheumatoid_arthritis	Ulcerative_colitis	0.4856
##			
##	HDL_cholesterol	LDL_cholesterol	0.48
##			
##	HDL_cholesterol	Platelet_counts	0.4609
##			
##	Platelet_counts	Triglycerides	0.4504
##			
##	LDL_cholesterol	Triglycerides	0.4151
##			
##	Creatinine_levels	Renal_function_related_traits_BUN	0.3915
##			
##	Psoriasis	Systemic_lupus_erythematosus	0.3911
##			
##	Renal_function_related_traits_BUN	Urate_levels	0.3689
##			
##	Alopecia_areata	C_reactive_protein	0.3686
##	-----		

The similarity dendrogram can be divided into separate groups:



```
## Cluster01 has    6 members
## Celiac_disease
## Multiple_sclerosis
## Kawasaki_disease
## Primary_biliary_cirrhosis
## Systemic_lupus_erythematosus
## Psoriasis
##
## Cluster02 has   14 members
## Type_2_diabetes
## Fasting_glucose_related_traits
## Red_blood_cell_traits
## Crohns_disease
## Migraine
## Systemic_sclerosis
## Ankylosing_spondylitis
## Platelet_counts
## Triglycerides
## HDL_cholesterol
## Vitiligo
## Progressive_supranuclear_palsy
## Liver_enzyme_levels_gamma_glutamyl_transferase
```

```

## LDL_cholesterol
##
## Cluster03 has 11 members
## Allergy
## Type_1_diabetes
## Primary_sclerosing_cholangitis
## Juvenile_idiopathic_arthritis
## Behcets_disease
## Ulcerative_colitis
## Rheumatoid_arthritis
## Autoimmune_thyroiditis
## Alopecia_areata
## C_reactive_protein
## Asthma
##
## Cluster04 has 8 members
## Bone_mineral_density
## Chronic_kidney_disease
## Alzheimers_combined
## Restless_legs_syndrome
## Atopic_dermatitis
## Urate_levels
## Renal_function_related_traits_BUN
## Creatinine_levels
##

```

The “Enrichment 1/2” columns show the average p-values of the group-specific SNPs-regulatory associations. A “-” sign indicates that an association is underrepresented. The “p-value” column shows whether the difference in the associations bwtween the groups is statistically significantly different.

```

## [1] "c1 vs. c2 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 45"
##
## -----
-----
##                               Row.names                               V2                               c1                               c2
##                               adj.P.Val
## -----
-----
## wgEncodeBroadHistoneGm12878H3k04me1StdPkV2  GM12878  H3K4me1  Histone Mods  6.263e-15  -0.010
15  3.557e-05
##
##                               by ChIP-seq Peaks from
##
##                               ENCODE/Broad
##
##
##  wgEncodeBroadHistoneGm12878H3k9me3StdPk  GM12878  H3K9me3  Histone Mods  2.129e-10  -0.0002
502  3.665e-05

```

##			by ChIP-seq Peaks from		
##			ENCODE/Broad		
##					
##	wgEncodeBroadHistoneGm12878H3k9acStdPk	GM12878 H3K9ac Histone Mods	3.849e-12	-0.20	
5	8.601e-07				
##			by ChIP-seq Peaks from		
##			ENCODE/Broad		
##					
##	wgEncodeBroadHistoneGm12878H3k27acStdPk	GM12878 H3K27ac Histone Mods	2.561e-11	-0.196	
7	0.0003686				
##			by ChIP-seq Peaks from		
##			ENCODE/Broad		
##					
##	wgEncodeBroadHistoneGm12878H2azStdPk	GM12878 H2A.Z Histone Mods by	8.379e-11	-0.070	
8	0.002582				
##			ChIP-seq Peaks from		
##			ENCODE/Broad		
##					
##	wgEncodeBroadHistoneCd20H3k04me2Pk	CD20+ H3K4me2 Histone Mods by	6.1e-09	-0.064	
18	3.557e-05				
##			ChIP-seq Peaks from		
##			ENCODE/Broad		
##					
##	wgEncodeBroadHistoneGm12878H3k4me2StdPk	GM12878 H3K4me2 Histone Mods	8.782e-09	-0.063	
92	8.601e-07				
##			by ChIP-seq Peaks from		
##			ENCODE/Broad		
##					
##	wgEncodeBroadHistoneCd20ro01794H3k27acPk	CD20+ R001794 H3K27ac Histone	1.057e-08	-0.098	
78	0.005575				
##			Mods by ChIP-seq Peaks from		
##			ENCODE/Broad		
##					

```

##          wgEncodeBroadHistoneCd20H2azPk          CD20+ H2A.Z Histone Mods by 5.608e-08 -0.033
72    0.000703
##
##
##
##
##
## wgEncodeBroadHistoneGm12878H3k04me3StdPkV2 GM12878 H3K4me3 Histone Mods 4.708e-08 -0.16
7    2.222e-05
##
##
##
##
## -----
-----
##
## [1] "c1 vs. c3 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 47"
##
## -----
-----
##          Row.names          V2          c1          c3
## adj.P.Val
## -----
-----
## wgEncodeBroadHistoneGm12878H3k04me1StdPkV2 GM12878 H3K4me1 Histone Mods 6.263e-15 0.2686
0.001041
##
##
##
##
##
##
##          wgEncodeBroadHistoneGm12878H3k9acStdPk          GM12878 H3K9ac Histone Mods 3.849e-12 0.2791
2.209e-05
##
##
##
##
##
##
##          wgEncodeBroadHistoneGm12878H3k27acStdPk          GM12878 H3K27ac Histone Mods 2.561e-11 0.1264
0.004196
##
##
##
##
##
##
##          wgEncodeBroadHistoneGm12878H2azStdPk          GM12878 H2A.Z Histone Mods by 8.379e-11 0.2377
0.01882

```

```
##                               ChIP-seq Peaks from

##                               ENCODE/Broad

##
##  wgEncodeBroadHistoneCd20ro01794H3k27acPk  CD20+  R001794  H3K27ac  Histone  1.057e-08  -0.4077
0.01648
##                               Mods by ChIP-seq Peaks from

##                               ENCODE/Broad

##
##  wgEncodeBroadHistoneGm12878H3k9me3StdPk  GM12878  H3K9me3  Histone  Mods   2.129e-10  0.03993
0.01648
##                               by ChIP-seq Peaks from

##                               ENCODE/Broad

##
##  wgEncodeBroadHistoneCd20H3k04me2Pk      CD20+  H3K4me2  Histone  Mods by   6.1e-09   0.7209
0.000875
##                               ChIP-seq Peaks from

##                               ENCODE/Broad

##
##  wgEncodeBroadHistoneGm12878H3k79me2StdPk  GM12878  H3K79me2  Histone  Mods  1.867e-08  -0.4995
7.28e-05
##                               by ChIP-seq Peaks from

##                               ENCODE/Broad

##
##  wgEncodeBroadHistoneGm12878H3k4me2StdPk  GM12878  H3K4me2  Histone  Mods   8.782e-09  0.5631
5.197e-05
##                               by ChIP-seq Peaks from

##                               ENCODE/Broad

##
##  wgEncodeBroadHistoneDnd41H3k04me1Pk      Dnd41  H3K4me1  Histone  Mods by  9.105e-08  -0.7878
0.0001171
##                               ChIP-seq Peaks from

##                               ENCODE/Broad

## -----
```

```

-----
##
## [1] "c1 vs. c4 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 58"
##
## -----
-----
##          Row.names          V2          c1          c4
##      adj.P.Val
## -----
--- -----
## wgEncodeBroadHistoneGm12878H3k04me1StdPkV2 GM12878 H3K4me1 Histone Mods  6.263e-15 -0.0086
89      0.0001299
##
##                      by ChIP-seq Peaks from
##
##                      ENCODE/Broad
##
##
## wgEncodeBroadHistoneGm12878H3k9me3StdPk  GM12878 H3K9me3 Histone Mods  2.129e-10 -1.413e
-05      5.165e-05
##
##                      by ChIP-seq Peaks from
##
##                      ENCODE/Broad
##
##
## wgEncodeBroadHistoneGm12878H3k27acStdPk  GM12878 H3K27ac Histone Mods  2.561e-11 -0.0070
23      0.0002497
##
##                      by ChIP-seq Peaks from
##
##                      ENCODE/Broad
##
##
## wgEncodeBroadHistoneGm12878H3k9acStdPk    GM12878 H3K9ac Histone Mods  3.849e-12  -0.138
3      4.143e-06
##
##                      by ChIP-seq Peaks from
##
##                      ENCODE/Broad
##
##
## wgEncodeBroadHistoneCd20ro01794H3k27acPk  CD20+ R001794 H3K27ac Histone 1.057e-08 -0.0010
05      0.001851
##
##                      Mods by ChIP-seq Peaks from
##
##                      ENCODE/Broad
##
##
##      wgEncodeBroadHistoneGm12878H2azStdPk  GM12878 H2A.Z Histone Mods by 8.379e-11  -0.16
1      0.008068

```

```

##                               ChIP-seq Peaks from

##                               ENCODE/Broad

##
##      wgEncodeBroadHistoneCd20H3k04me2Pk      CD20+ H3K4me2 Histone Mods by 6.1e-09      -0.015
43      6.552e-05

##                               ChIP-seq Peaks from

##                               ENCODE/Broad

##
##      wgEncodeBroadHistoneGm12878H3k79me2StdPk  GM12878 H3K79me2 Histone Mods 1.867e-08 -0.0056
56      5.047e-06

##                               by ChIP-seq Peaks from

##                               ENCODE/Broad

##
##      wgEncodeBroadHistoneGm12878H3k4me2StdPk  GM12878 H3K4me2 Histone Mods 8.782e-09 -0.028
11      4.143e-06

##                               by ChIP-seq Peaks from

##                               ENCODE/Broad

##
##      wgEncodeBroadHistoneGm12878H3k04me3StdPkV2 GM12878 H3K4me3 Histone Mods 4.708e-08 -0.016
21      2.43e-05

##                               by ChIP-seq Peaks from

##                               ENCODE/Broad

## -----
-----
##
## [1] "c2 vs. c3 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 32"
##
## -----
-----
##                               Row.names                               V2                               c2                               c3
      adj.P.Val
## -----
-- -----
##      wgEncodeBroadHistoneK562H3k36me3StdPk      K562 H3K36me3 Histone Mods by 0.001587      -0.0097
93      0.03505

##                               ChIP-seq Peaks from

```



##		ENCODE/Broad	
##			
##	wgEncodeBroadHistoneNhekH4k20me1StdPk	NHEK H4K20me1 Histone Mods by	0.0009324 -0.036
4	0.06562		
##		ChIP-seq Peaks from	
##		ENCODE/Broad	
##			
##	wgEncodeBroadHistoneNhdfadH4k20me1Pk	NHDF-Ad H4K20me1 Histone Mods	0.001681 -0.0273
7	0.04472		
##		by ChIP-seq Peaks from	
##		ENCODE/Broad	
##			
##	wgEncodeBroadHistoneNhaH3k79me2Pk	NH-A H3K79me2 Histone Mods by	0.01045 -0.0129
4	0.01579		
##		ChIP-seq Peaks from	
##		ENCODE/Broad	
##			
##	wgEncodeBroadHistoneA549H3k79me2Dex100nmPk	A549 DEX 100 nM H3K79me2	0.008678 -0.0211
5	0.01579		
##		Histone Mods by ChIP-seq	
##		Peaks from ENCODE/Broad	
##			
##	wgEncodeBroadHistoneNhlfH3k79me2Pk	NHLF H3K79me2 Histone Mods by	0.005332 -0.0517
9	0.03198		
##		ChIP-seq Peaks from	
##		ENCODE/Broad	
##			
##	wgEncodeBroadHistoneNhdfadH3k79me2Pk	NHDF-Ad H3K79me2 Histone Mods	0.01755 -0.0217
8	0.03732		
##		by ChIP-seq Peaks from	
##		ENCODE/Broad	
##			
##	wgEncodeBroadHistoneA549H3k36me3Etoh02Pk	A549 EtOH 0.02% H3K36me3	0.01319 -0.0312
9	0.02689		

```

##                               Histone Mods by ChIP-seq

##                               Peaks from ENCODE/Broad

##
##   wgEncodeBroadHistoneOsteoH3k79me2Pk   Osteoblasts H3K79me2 Histone   0.02111  -0.0199
3   0.01579
##                               Mods by ChIP-seq Peaks from

##                               ENCODE/Broad

##
##   wgEncodeBroadHistoneA549H3k79me2Etoh02Pk   A549 EtOH 0.02% H3K79me2   0.01245  -0.0361
2   0.02647
##                               Histone Mods by ChIP-seq

##                               Peaks from ENCODE/Broad

## -----
-----
##
## [1] "c2 vs. c4 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 9"
##
## -----
-----
##                               Row.names                               V2                               c2                               c4
##   adj.P.Val
## -----
-----
##   wgEncodeBroadHistoneHmecH3k36me3StdPk   HMEC H3K36me3 Histone Mods by   0.01773  -0.000284
7   0.04273
##                               ChIP-seq Peaks from

##                               ENCODE/Broad

##
##   wgEncodeBroadHistoneK562H3k36me3StdPk   K562 H3K36me3 Histone Mods by 0.001587  -0.00739
0.1287
##                               ChIP-seq Peaks from

##                               ENCODE/Broad

##
##   wgEncodeBroadHistoneNhekH4k20me1StdPk   NHEK H4K20me1 Histone Mods by 0.0009324  -0.05835
0.198
##                               ChIP-seq Peaks from

```

##		ENCODE/Broad		
##				
##	wgEncodeBroadHistoneK562Nsd2ab75359Pk	K562 NSD2 (ab75359) Histone	0.001909	-0.03803
	0.1947			
##		Mods by ChIP-seq Peaks from		
##		ENCODE/Broad		
##				
##	wgEncodeBroadHistoneK562RestPk	K562 REST Histone Mods by	0.007983	-0.01042
	0.1487			
##		ChIP-seq Peaks from		
##		ENCODE/Broad		
##				
##	wgEncodeBroadHistoneNhdfadH4k20me1Pk	NHDF-Ad H4K20me1 Histone Mods	0.001681	-0.1255
	0.2067			
##		by ChIP-seq Peaks from		
##		ENCODE/Broad		
##				
##	wgEncodeBroadHistoneK562NcorPk	K562 NCoR Histone Mods by	0.04478	-0.005174
	0.0846			
##		ChIP-seq Peaks from		
##		ENCODE/Broad		
##				
##	wgEncodeBroadHistoneGm12878H3k36me3StdPk	GM12878 H3K36me3 Histone Mods	0.2219	-0.001063
	0.05838			
##		by ChIP-seq Peaks from		
##		ENCODE/Broad		
##				
##	wgEncodeBroadHistoneNhekH3k9me1StdPk	NHEK H3K9me1 Histone Mods by	0.2002	-0.001186
	0.02253			
##		ChIP-seq Peaks from		
##		ENCODE/Broad		
##				
##	wgEncodeBroadHistoneHepg2H3k36me3StdPk	HepG2 H3K36me3 Histone Mods	0.03916	-0.006255
	0.1623			

[illegible]

```

##                                     ENCODE/Broad

##
##  wgEncodeBroadHistoneOsteoH3k27me3Pk  Osteoblasts H3K27me3 Histone  0.0006062  -0.1426
0.6051
##                                     Mods by ChIP-seq Peaks from

##                                     ENCODE/Broad

##
##  wgEncodeBroadHistoneNhaH3k27me3StdPk  NH-A H3K27me3 Histone Mods by 0.0001926  -0.9361
0.6051
##                                     ChIP-seq Peaks from

##                                     ENCODE/Broad

##
##  wgEncodeBroadHistoneOsteoP300kat3bPk  Osteoblasts P300 KAT3B      0.0004775  -0.5556
0.6051
##                                     Histone Mods by ChIP-seq

##                                     Peaks from ENCODE/Broad

##
##  wgEncodeBroadHistoneGm12878Ezh239875Pk  GM12878 EZH2 (39875) Histone  0.02094  -0.01915
0.6051
##                                     Mods by ChIP-seq Peaks from

##                                     ENCODE/Broad

##
##  wgEncodeBroadHistoneHepg2H3k27me3StdPk  HepG2 H3K27me3 Histone Mods  0.0008383  -0.6783
0.6051
##                                     by ChIP-seq Peaks from

##                                     ENCODE/Broad

## -----
-----

```

**Text mining question 2:** Are the terms associated stronger with the diseases in one vs. the other cluster based on the literature strength? Are the terms themselves related based on the literature? *Expected answer:* Yes, the literature associations should confirm the relationships.

# Summary

1. Again, cluster 1 is strongly distinct. Cluster 2 is less so. Histone marks seem all active.

C1	C2	C3	C4
C1	Cell types: Gm12878, CD20+ Reg: H3K4me1, H3K9me3, H3K9ac, H3K27ac, H2az, H3K4me2	Cell types: Gm12878, CD20+ Reg: H3K4me1, H3K9me3, H3K9ac, H3K27ac, H2az, H3K4me2	Cell types: Gm12878, CD20+ Reg: H3K4me1, H3K9me3, H3K9ac, H3K27ac, H2az, H3K4me2
C2		Cell types: K562, NHEK, NHDF-Ad, NH-A, HMEC Reg: H3K36me3, H4K20me1, H3K79me2	Nothing significant
C3			Nothing significant
C4			

# Analysis of all regulatory datasets

Out of all regulatory datasets, we select all. The goal here is to get potentially tighter clustering.

## [ 1 ] 4498 39
## [ 1 ] 2969 39



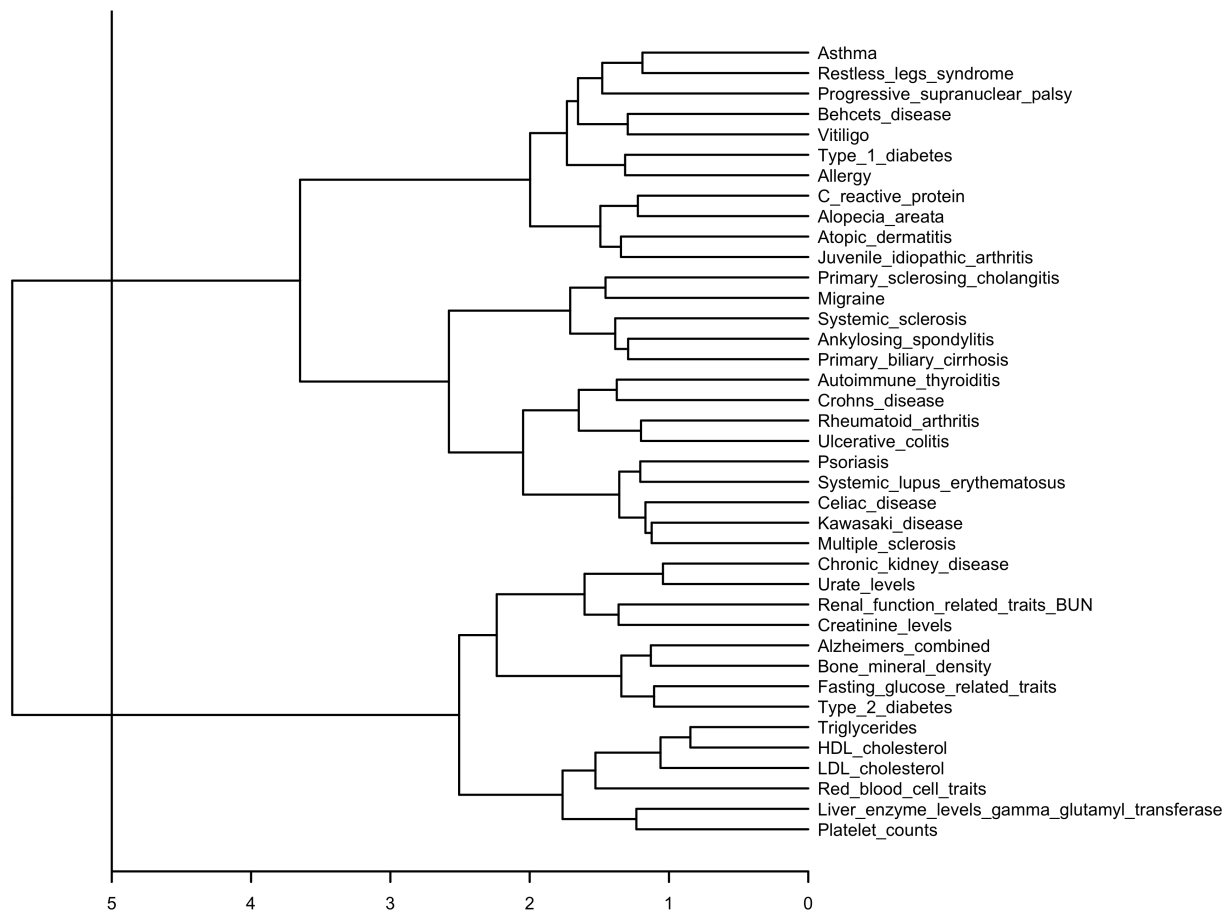
```

##
## -----
--
##          Disease 1                      Disease 2          Corr. coefficient
t
## -----
--
##          HDL_cholesterol                Triglycerides        0.473
##
##          LDL_cholesterol                Triglycerides        0.4314
##
##          Chronic_kidney_disease          Urate_levels          0.3742
##
##          HDL_cholesterol                LDL_cholesterol        0.3475
##
##          Bone_mineral_density            Type_2_diabetes        0.3225
##
##          Multiple_sclerosis              Primary_biliary_cirrhosis 0.316
##
##          Alzheimers_combined             Type_2_diabetes        0.2999
##
## Liver_enzyme_levels_gamma_glutamyl_transferase  Urate_levels          0.2976
##
##          Fasting_glucose_related_traits  Type_2_diabetes        0.2972
##
## Liver_enzyme_levels_gamma_glutamyl_transferase  Platelet_counts        0.2944
## -----
--

```

The similarity dendrogram can be divided into separate groups:





```

## Cluster01 has 14 members
## Platelet_counts
## Liver_enzyme_levels_gamma_glutamyl_transferase
## Red_blood_cell_traits
## LDL_cholesterol
## HDL_cholesterol
## Triglycerides
## Type_2_diabetes
## Fasting_glucose_related_traits
## Bone_mineral_density
## Alzheimers_combined
## Creatinine_levels
## Renal_function_related_traits_BUN
## Urate_levels
## Chronic_kidney_disease
##
## Cluster02 has 25 members
## Multiple_sclerosis
## Kawasaki_disease
## Celiac_disease
## Systemic_lupus_erythematosus
## Psoriasis
## Ulcerative_colitis
## Rheumatoid_arthritis
## Crohns_disease
## Autoimmune_thyroiditis
## Primary_biliary_cirrhosis
## Ankylosing_spondylitis
## Systemic_sclerosis
## Migraine
## Primary_sclerosing_cholangitis
## Juvenile_idiopathic_arthritis
## Atopic_dermatitis
## Alopecia_areata
## C_reactive_protein
## Allergy
## Type_1_diabetes
## Vitiligo
## Behcets_disease
## Progressive_supranuclear_palsy
## Restless_legs_syndrome
## Asthma
##

```

The “Enrichment 1/2” columns show the average p-values of the group-specific SNPs-regulatory associations. A “-” sign indicates that an association is underrepresented. The “p-value” column shows whether the difference in the associations bwtween the groups is statistically significantly different.

## [1] "c1 vs. c2 , number of degs significant at adj.p.val<0.5 and 2-fold diff: 0"

##

## -----

## Row.names V2 c1

c2 adj.P.Val

## -----

## wgEncodeHaibTfbsMcf7Hdac2sc6296V0422111PkRep2 MCF-7 HDAC2 v042211.1 1.561e-07

0.8353 0.4918

## ChIP-seq Peaks Rep 2 from

## ENCODE/HAIB

##

## wgEncodeHaibTfbsA549Tead4sc101184V0422111PkRep1 A549 TEAD4 v042211.1 ChIP-seq 2.373e-06

0.7976 0.4968

## Peaks Rep 1 from ENCODE/HAIB

##

## wgEncodeSydhTfbsH1hescSuz12UcdPk H1-hESC SUZ12 UC Davis 1.619e-05

0.6341 0.4822

## ChIP-seq Peaks from

## ENCODE/SYDH

##

## wgEncodeSydhTfbsMcf7Gata3sc269UcdPk MCF-7 GATA3 SC269 UC Davis 4.078e-05

0.7946 0.5005

## ChIP-seq Peaks from

## ENCODE/SYDH

##

## wgEncodeGisChiaPetMcf7EraaInteractionsRep2 MCF-7 ERalpha a ChIA-PET 0.0008577

-0.1465 0.1052

## Interactions Rep 2 from

## ENCODE/GIS-Ruan

##

## wgEncodeBroadHistoneNhekH4k20me1StdPk NHEK H4K20me1 Histone Mods by 0.003353

-0.0967 0.1683

## ChIP-seq Peaks from

## ENCODE/Broad

```

##
## wgEncodeCshlLongRnaSeqK562ChromatinTotalJunctions K562 chromatin total RNA-seq 0.02711
-0.01382 0.3846
##
##
##
##
##
## wgEncodeGisRnaPetK562NucleusPapClustersRep1 K562 nucleus polyA+ 0.005089
-0.07474 0.1233
##
## clone-based RNA PET Clusters
##
## Rep 1 from ENCODE/GIS
##
##
## wgEncodeGisRnaPetK562ChromatinTotalClustersRep1 K562 chromatin total 0.01422
-0.03698 0.1849
##
## clone-based RNA PET Clusters
##
## Rep 1 from ENCODE/GIS
##
##
## wgEncodeGisRnaPetHepg2CytosolPapClustersRep1 HepG2 cytosol polyA+ 0.006008
-0.1193 0.1068
##
## clone-based RNA PET Clusters
##
## Rep 1 from ENCODE/GIS
##
## -----
-----

```

# Summary

The picture is not as good as when we are taking subsets of regulatory datasets.