eQTL analysis of gene CDA and its significance based on cancer

Soongsil Univ. School of Systemic Biomedical Science
BakChaeRin 20210387

1. Introduction

A gene called CDA or CDD located in chr1 is associated with toxicity to anti-tumor drugs because CDA proteins metabolize them.[1 2] In certain leukemia, there is a possibility of treating leukemia through apoptosis by silencing CDA.[3] And in some cancers, patients with low CDA expression have a good prognosis.[4] However, although there have been many attempts to analyze CDA in terms of the enzyme function, little research has been done on how mRNA expression is regulated by the influence of genetic factors. Therefore, I performed a GWAS analysis to find the eQTL that affects CDA expression in order to find out the effect of genetic factors on CDA which is heavily related to cancer.

2. Methods

Analysis was performed using mRNA sequencing expression data obtained from the European lymphoblast cell line. Plink was used as a tool.[5] After performing line regression with an additional model to determine which snp is related to the expression level of CDA, eQTL was obtained by applying a threshold to the p-value. For the found signals, haploview was executed to check LD blocks, and each block was treated as eQTL.

To find out whether there are more phenotypes related to the selected eQTLs, linear regression was performed with plink to compare the p-value between the phenotypes from chr1 to 22 and eQTL, and in this case, the threshold was applied as 4.89*10^-6 divided by the number of genes.

To estimate the function of the found eQTL, ncbi, uniprot, regulomedb, haploreg, ucsc(genehancer, refTSS, ENCODE cCREs, FANTOM5, Genecard, JASPAR CORE 2022 transcription factor binding sites), 3D Genome Browser(ChiA-PET visualized data), DNA binding site predictor for Cys2His2 zinc finger proteins, ensemblensembl regulatory build, FANTOM5), SwissRegulon(EPDnew), and JASPAR2024 were used.

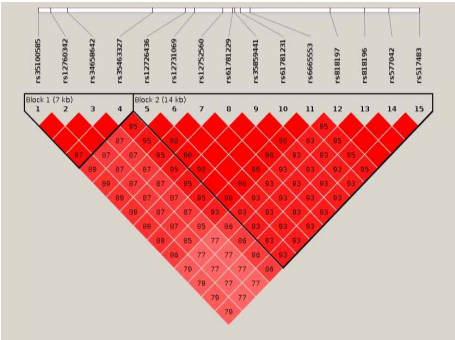
3. Results

snp	position	gene	allele	MAF	beta	p
rs35100585	1:20906232	Intergenic (9.2kb 5' of CDA)	G/A	0.07181	0.2310	4.17E-09
rs12760342	1:20907048	Intergenic (8.4kb 5' of CDA)	A/C	0.07314	0.2326	2.46E-09
rs34658642	1:20908101	Intergenic (7.3kb 5' of CDA)	T/C	0.07314	0.2326	2.46E-09
rs35463327	1:20913374	Intergenic (2.1kb 5' of CDA)	A/G	0.06782	0.2577	2.02E-10
rs12726436	1:20915441	CDA : 2KB Upstream Variant	A/G	0.10900	0.2718	4.92E-16
rs12731069	1:20916012	CDA : Intron Variant	A/G	0.11040	0.2626	7.24E-15
rs12752560	1:20917765	CDA : Intron Variant	T/C	0.11040	0.2582	1.28E-14
rs61781229	1:20918398	CDA : Intron Variant	A/G	0.11040	0.2582	1.28E-14
rs35859441	1:20918526	CDA : Intron Variant	C/G	0.11040	0.2582	1.28E-14
rs61781231	1:20918862	CDA : Intron Variant	G/A	0.11170	0.2673	9.06E-16
rs6665553	1:20919470	CDA : Intron Variant	T/C	0.11040	0.2611	6.12E-15
rs818197	1:20924483	CDA : Intron Variant	A/C	0.17420	0.1557	1.18E-08
rs818196	1:20926631	CDA : Intron Variant	G/C	0.17290	0.1587	8.78E-09
rs577042	1:20928154	CDA : Intron Variant	C/T	0.17290	0.1587	8.78E-09
rs517483	1:20930035	CDA : Intron Variant	C/T	0.17290	0.1587	8.78E-09
rs6755057	2:25658098	DTNB : Intron Variant LOC124900608 : Non Coding Transcript Variant	G/C	0.08511	0.2027	2.10E-08
rs10865350	2:25658384	DTNB : Intron Variant LOC124900608 : Non Coding Transcript Variant	T/C	0.08245	0.2063	1.68E-08
rs7566839	2:25670813	DTNB : Intron Variant	T/C	0.08245	0.2063	1.68E-08
rs6715574	2:25676745	DTNB : Intron Variant	A/G	0.07713	0.2199	7.52E-09
rs116109937	5:144125259	Intergenic (210kb 5' of 7SK)	T/C	0.07048	0.2323	1.45E-08

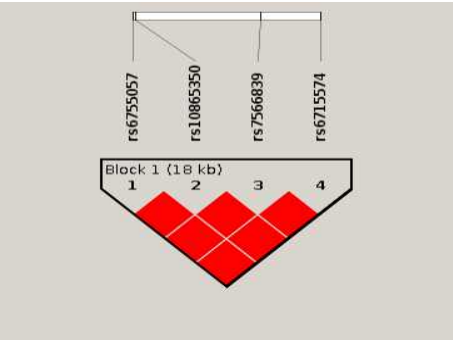
<Table 1> The eQTLs found through association

: LD blocks were classified by lines, the position was written based on hg19, and minor/major allele marked.

When the p-value threshold was applied, a total of 20 signals were found as shown in <Table 1>.



<Figure 1> LD block result in chr1



<Figure 2> LD block result in chr2

Through Haploview, it was confirmed that 20 signals formed a total of four LD blocks: two from chr1, one from chr2, and one from chr5 (using standard TDT, ignore pairwise comparisons of markers > 500kb apart).

4. Discussion

snp	position	RegulomeDB	HaploReg	GeneHancer	GeneCards
rs35100585	1:20906232	transcription, polycomb, heterochromatin	-	-	-
rs12760342	1:20907048	transcription, polycomb, heterochromatin	promoter	-	-
rs34658642	1:20908101	transcription, polycomb, heterochromatin	promoter	-	-
rs35463327	1:20913374	transcription, polycomb	enhancer, promoter	-	-
rs12726436	1:20915441	TSS, enhancer, polycomb	TSS, enhancer, promoter	promoter, enhancer	TSS
rs12731069	1:20916012	TSS, enhancer, polycomb	TSS, transcription, enhancer, promoter	promoter, enhancer	-
rs12752560	1:20917765	TSS, enhancer, polycomb	TSS, transcription, enhancer, promoter	promoter, enhancer	-
rs61781229	1:20918398	TSS, transcription, enhancer, polycomb	TSS, transcription, enhancer, promoter	-	-
rs35859441	1:20918526	transcription, enhancer, polycomb	TSS, transcription, enhancer, promoter	-	-
rs61781231	1:20918862	TSS, transcription, enhancer, polycomb	transcription, enhancer, promoter	-	-
rs6665553	1:20919470	enhancer, polycomb	transcription, enhancer, promoter	-	-
rs818197	1:20924483	transcription, enhancer, polycomb	TSS, transcription, enhancer, promoter	enhancer	-
rs818196	1:20926631	TSS, transcription, enhancer, polycomb	TSS, transcription, enhancer, promoter	-	-
rs577042	1:20928154	transcription, enhancer, polycomb	transcription, enhancer, promoter	enhancer	-
rs517483	1:20930035	TSS, transcription, enhancer, polycomb	TSS, enhancer, promoter, DNase	enhancer	-
rs6755057	2:25658098	transcription, heterochromatin	enhancer	-	-
rs10865350	2:25658384	transcription, ZNF, polycomb, heterochromatin	enhancer	-	-
rs7566839	2:25670813	transcription, polycomb	enhancer, promoter	-	-
rs6715574	2:25676745	transcription, enhancer, polycomb	TSS, enhancer, promoter	-	-
rs116109937	5:144125259	heterochromatin	-	-	-

<Table 2> The predicted functions of each snp examined from the database : regulomedb, haploreg, genehancer, and genecards were investigated and organized in a table. None or miss data was treated as -, and each LD block was divided by a thick line.

The second eQTL of chr1 was found to be very likely to function as a TSS or promoter or enhancer through the support data in <Table 2>.

In addition, as a result of checking the ChiA-PET data for all found snp belonging to the second LD block (hg19, K562 leukemia cell line, GM12878 lymphoblast cell line, using snyder and ruan lab data, chromatin interaction was confirmed in the range of 500 kb), it was confirmed that chromatin interaction occurred with the genes MUL1, FAM43B, CDA, PINK1, DDOST, and KIF17. Therefore, when the second LD block functions as a TSS or promoter or enhancer, it was suggested that the target might be MUL1, FAM43B, CDA, PINK1, DDOST, and KIF17. As a result of performing linear association on the above gene with plink, PINK1 and KIF17 did not satisfy the p-value<0.05, but CDA satisfies the p-value. According to gene targets for GeneHancer data from genecard, CDA had the highest score.

Judging from the above two results, the second eQTL of chr1 can function as a TSS or promoter or an enhancer, and its target may be assumed to be CDA.

Also, further investigation into the function of rs12726436 follows:
According to the trascript of ensembl and refTSS, rs12726436 is the TSS. Since rs12726436 is located in the 2KB Upstream direction from the CDA, it is very likely that it is a promoter. According to ENCODE cCREs and SwissRegulon(EPDnew), rs12726436 will belong to the promoter. rs12726436 will belong to the promoter because it was confirmed that POLR2A, TF(NRF1) is bound through regulomedb and haploreg's K562 leukemia cell line, and immune cell target chip-seq results.

According to JASPAR CORE 2022 Transcription Factor Binding Sites, several TFs (SCRT1, Zic3, Zic2, Zic1::Zic2, ZNF331) will be bound and thus rs12726436 will be a promoter. According to ensemble(ensemble regulatory build, FANTOM5), rs12726436 is predicted to be a promoter because TF(MYBL1:MAX, SRF) will be bound.

I investigated the motif of TFs that bind to rs12726436 or were found to be possible (using JASPAR2024) and compared it with the major/minor allele of rs12726436. As a result, NRF1, SCRT1, Zic3, Zic2, Zic1::Zic2, ZNF331, and SRF bind more to DNA when it is major allele. However, the MYBL1::MAX TF complex is more likely to bind to DNA when rs12726436 is minor allele(T) than when it is major allele(C). That is, the MYBL1:MAX TF complex would bind better to the promoter when rs12726436 is minor allele.

Therefore, since MYBL1::MAX upregulates the promoter of CDA, it is expected that the expression level of CDA would be higher when rs12726436 is a minor allele than when it is a major allele (beta>0). In fact, since it is not known what function the MYBL1::MAX TF complex performs, further research on this will be able to find out a more detailed regulatory action process.

As a result, among snp belonging to eQTL, rs12726436 is expected to belong to the promoter of CDA, and it is also possible that it is the TSS of CDA within the promoter. And since MYBL1::MAX can more strongly bind to the promoter of CDA when it is the minor allele, the expression of CDA increases as the number of minor alleles increases by MYBL1::MAX which upregulates the expression of CDA.

Through this study, it was possible to select eQTLs that affect the expression level of CDA, and for one snp of them, the cause is suggested. People with low CDA activity are more likely to experience severe/lethal toxicity from solid tumor drugs.[1] [2] And in some cancers, tumor patients with little or no CDA expression have a better prognosis.[4] Therefore, it is possible to estimate the expression level of CDA by comparing the patient's genome and eQTL, and the degree of resistance to anti-tumor drugs of solid tumor and its prognosis can be predicted in advance through the expression level of CDA.

Although not mainly covered in the text, the CLTB gene of chr5 also had associations with the discovered eQTLs. It was found that the expression level of both CDA and CLTB was significantly changed by rs116109937 present in chr5 (CLTB: p=1.723e-06). As a result of confirming the chromatin interaction of rs116109937, it was confirmed that the interaction occurred very widely within chr5. As a result of confirmation by chia-pet, the chromatin interaction targeting RAD21 in the GM12878 cell line was confirmed from chr5:143583276-143583772 to chr5:144617287-144617840 (hg19). As a result of confirmation by hichip, it was confirmed that the interaction targeting cohesin in GM12878 was

chr5:1435800-143590000 to chr5:144610000-144620000 (hg19). Although there are not many genes within the corresponding range of interactions, further studies are needed on how rs116109937 can broadly regulate chr5.

In addition, among the eQTLs of CDA, the eQTL included in the DTNB intron of chr2 also has the possibility of zinc finger. I tried to find out whether the target of zinc finger is CDA through a tool, but no significant result was obtained. However, since DTNB may be related to the expression of CDA, further research on this will be needed.

5. Conclusion

The eQTL of chr1 to which rs12726436, rs12731069, rs12752560, rs61781229, rs35859441, rs61781231, rs6665553, rs818197, rs818196, rs577042, and rs517483 belongs may function as TSS or promoter, and its target is expected to be CDA.

In particular, since rs12726436 is expected to function as both a TSS and a promoter of CDA, it is expected that the CDA expression level increases as the number of minor alleles(A) increases. When it is a minor allele, MYBL1::MAX may more strongly bind to the promoter of CDA and might upregulates CDA, so that the CDA expression level increases as the number of minor alleles increases.

6. References

1) Fanciullino R, Farnault L, Donnette M. CDA as a predictive marker for life-threatening toxicities in patients with AML treated with cytarabine. Blood Adv. 2018 Mar 13;2(5):462-469. doi: 10.1182/bloodadvances.2017014126. Erratum in: Blood Adv. 2018 May 8;2(9):1013. PMID: 29490977; PMCID: PMC5851420.

2) Iyer SN, Ankala A, Singhal RS. Determination of common genetic variants in cytidine deaminase (CDA) gene in Indian ethnic population. Gene. 2013 Jul 15;524(1):35-9. doi: 10.1016/j.gene.2013.04.015. Epub 2013 Apr 21. PMID: 23612254.

3) Wei XF, Feng YF, Chen QL. CDA gene silencing regulated the proliferation and apoptosis of chronic myeloid leukemia K562 cells. Cancer Cell Int. 2018 Jul 9;18:96. doi: 10.1186/s12935-018-0587-y. PMID: 30002603; PMCID: PMC6038203.

4) Mameri H, Buhagiar-Labarchède G, Fontaine G. Cytidine deaminase deficiency in tumor cells is associated with sensitivity to a naphthol derivative and a decrease in oncometabolite levels. Cell Mol Life Sci. 2022 Aug 4;79(8):465. doi: 10.1007/s00018-022-04487-9. PMID: 35925417; PMCID: PMC9352748.

5) Purcell S, Neale B, Todd-Brown K. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559-75. doi: 10.1086/519795. Epub 2007 Jul 25. PMID: 17701901; PMCID: PMC1950838.