

유전자 CDA에 대한 eQTL 분석과 암을 기준으로 한 의의

의생명시스템학부 20210387 박채린

1. Introduction

CDA또는 CDD라 불리는 유전자는 단백질이 항종양제를 대사하므로 약물에 대한 독성과 관련이 있고^{1, 2}, 특정 백혈병에서는 CDA를 silencing함으로써 세포사멸을 통해 백혈병을 치료할 가능성이 있으며³, 일부 암의 경우에는 CDA 발현이 적은 환자가 예후가 좋은⁴ 등 암과 다양하게 관련이 있다. 하지만 CDA를 mRNA와 단백질의 enzyme 작용 관점에서 분석하고자 하는 시도는 많았지만, 유전적 요인이 미치는 영향에 의해 어떻게 mRNA 발현량이 조절되는지에 대해서는 거의 연구된 바가 없다. 따라서 나는 암과 많은 관련이 있는 CDA에게 유전적 요인이 미치는 영향을 알아보기 위해 CDA 발현량에 영향을 미치는 eQTL을 찾고자 GWAS 분석을 수행하였다.

2. Methods

유러피안의 림포블라스트 cell line에서부터 얻은, mRNA sequencing 발현 데이터를 이용하여 분석을 수행하였다. tool로는 plink⁵를 이용하였다. chr1번에 위치하는 CDA에 대하여, 어떤 snp가 발현량과 연관있는지, additive model로 linear regression을 수행한 뒤, 위의 p-value에 대해 threshold를 적용하여 eQTL을 얻었다. 발견된 signals에 대해 haploview를 실행하여 LD block을 확인하였고 각 block을 eQTL로 취급하였다.

선정된 eQTL들과 연관있는 phenotype이 더 있는지, plink로 linear regression을 수행해 chr1번부터 22번까지의 phenotype들과 eQTL간의 p-value를 비교하였으며, 이때는 threshold를 유전자 수로 나눈 4.89×10^{-6} 으로 적용하였다.

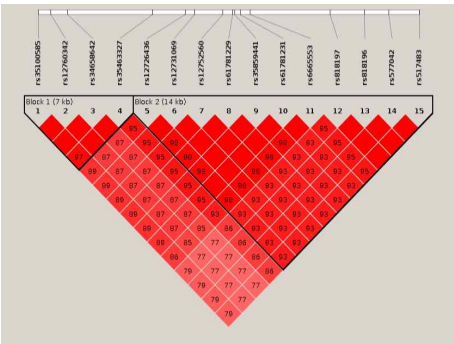
발견된 eQTL의 기능을 추측하는 데에는 ncbi, uniprot, regulomedb, haploreg, ucsc(genehancer, refTSS, ENCODE cCREs, FANTOM5, genecard, JASPAR CORE 2022 Transcription Factor Binding Sites), 3D Genome Browser(ChiA-PET visualized data), DNA binding site predictor for Cys2His2 Zinc Finger Proteins, ensembl(ensembl regulatory build, FANTOM5), SwissRegulon(EPDnew), JASPAR2024를 이용하였다.

3. Results

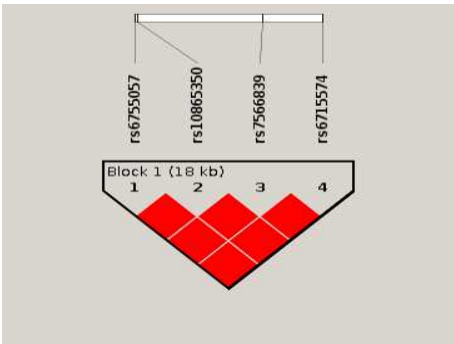
snp	position	gene	allele	MAF	beta	p
rs35100585	1:20906232	Intergenic (9.2kb 5' of CDA)	G/A	0.07181	0.2310	4.17E-09
rs12760342	1:20907048	Intergenic (8.4kb 5' of CDA)	A/C	0.07314	0.2326	2.46E-09
rs34658642	1:20908101	Intergenic (7.3kb 5' of CDA)	T/C	0.07314	0.2326	2.46E-09
rs35463327	1:20913374	Intergenic (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	G/C	0.08511	0.2027	2.10E-08
		LOC124900608 : Non Coding Transcript Variant				
rs10865350	2:25658384	DTNB : Intron Variant	T/C	0.08245	0.2063	1.68E-08
		LOC124900608 : Non Coding Transcript Variant				
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

<table1> association을 통해 발견한 eQTL : 실선으로 LD block을 구분하였다, position은 hg19를 기준으로 작성되었다, allele는 minor/major allele를 표시하였다.

p-value threshold를 적용하였을 때, <표1>과 같이 총 20개의 signal이 발견되었다.



<figure1> chr1에서의 LD block 확인 결과



<figure2> chr2에서의 LD block 확인 결과

haploview를 통해 20개의 signal이 chr1에서 두 개, chr2에서 한 개, chr5에서 한 개, 총 4개의 LD block을 형성함을 확인하였다(standard TDT 사용, ignore pairwise comparisons of markers > 500kb apart).

4. Discussion

1. chr1에 속한 두 번째 eQTL의 기능 예측

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	-	-	-

<table2> 데이터베이스로부터 조사한 각 snp의 예측되는 기능 : regulomedb, haploreg, genehancer, genecards에서 snp가 속한 구역의 기능을 조사하여 표로 정리하였음, none 또는 miss는 -로 처리함, 굵은 선으로 각 LD block을 구분하였음.

chr1의 두 번째 eQTL은 <표2>의 뒷받침데이터를 통해 TSS 또는 promoter 또는 enhancer로 기능할 가능성이 매우 높을 것이라는 결과가 도출되었다.

또한 두 번째 LD block에 속한 발견된 모든 snp에 대하여 ChiA-PET 데이터를 확인한 결과 (hg19, K562 백혈병 세포 cell line, GM12878 lymphoblast cell line, snyder와 ruan lab data를 사용, 500kb범위에서 chromatin interaction 확인) , MUL1, FAM43B, CDA, PINK1, DDOST, KIF17 유전자들과 chromatin interaction이 일어남을 확인할 수 있었다. 따라서 두 번째 LD block이 TSS 또는 promoter 또는 enhancer로서 기능할 때, 그 타겟은 MUL1, FAM43B, CDA, PINK1, DDOST, KIF17 중에 있을 수 있다는 가능성이 제시되었다.

plink로 위 유전자에 대해 linear association을 수행해본 결과, PINK1과 KIF17은 p-value<0.05를 만족시키지 못하였지만, CDA는 p-value를 충족시킨다.

genecard의 Gene targets for GeneHancer 데이터에 따르면 가장 score가 높은 것은 CDA였다.

위 두 결과를 종합하면, chr1의 두 번째 eQTL은 TSS 또는 promoter 또는 enhancer로 기능할 수 있는데, 그 타겟은 CDA라 추측해볼 수가 있다.

또한 rs12726436의 기능에 대해 추가적으로 조사한 바에 따르면 다음과 같다. :

refTSS, ensembl의 trascript에 따르면 rs12726436는 TSS이다. rs12726436는 CDA로부터 2KB Upstream 방향에 위치하기 때문에 promoter일 가능성이 매우 높다. ENCODE cCREs, SwissRegulon(EPDnew)에 따르면 rs12726436은 promoter에 속할 것이다.

rs12726436에는 regulomedb, haploreg의 K562 백혈병 세포 cell line, 면역세포 대상 chip-seq 결과를 통해, POLR2A, TF(NRF1)가 bind함을 확인하였으므로 rs12726436이 promoter에 속할 것이다. JASPAR CORE 2022 Transcription Factor Binding Sites에 따르면 여러 TF(SCRT1, Zic3, Zic2, Zic1::Zic2, ZNF331)가 bind할 것으로 예측되어 promoter일 것으로 예측된다. ensemble(ensemble regulatory build, FANTOM5)에 따르면 rs12726436은 TF(MYBL1::MAX, SRF)가 bind할 것이므로 따라서 promoter로 예측된다.

rs12726436과 bind하거나 또는 그 가능성이 있다고 나온 TF들의 motif를 조사하여 rs12726436의 major/minor allele와 비교해보았다(JASPAR2024 사용). 그 결과, NRF1, SCRT1, Zic3, Zic2, Zic1::Zic2, ZNF331, SRF는 major allele일 때 DNA와 더 많이 bind하였다. 그러나 MYBL1::MAX TF complex는 rs12726436이 major allele(C)일 때보다 minor allele(T)일 때 DNA에 bind하는 경우가 많았다. 즉 MYBL1::MAX TF complex는 rs12726436이 minor allele일 때 promoter에 더 잘 bind할 것이다.

따라서 MYBL1::MAX가 CDA의 promoter를 upregulation시키기 때문에 rs12726436가 major allele를 가질 때보다 minor allele일 때 CDA의 발현량이 더 높을 것(beta>0)이라고 예상하였다. 실제로 MYBL1::MAX TF complex는 어떤 기능을 수행하는지 밝혀져있지 않기 때문에, 이에 대한 추가적인 연구가 이루어지면 보다 자세한 조절 작용 과정을 알아낼 수 있을 것이다.

결과적으로 eQTL에 속하는 snp 중 rs12726436은 CDA의 promoter에 속하는 것으로 예상되며, promoter 내에서도 CDA의 TSS일 가능성이 있다. 그리고 minor allele일 때 MYBL1::MAX가 CDA의 promoter에 더 강하게 bind할 수 있으므로, MYBL1::MAX가 CDA의 발현을 upregulation시켜 minor allele의 수가 증가할수록 CDA 발현량이 증가하는 것일 것이다.

이번 연구를 통해 CDA의 발현량에 영향을 미치는 eQTL들을 선별할 수 있었고, 그 중 하나의 snp에 대해서는 그 원인까지 추측해보았다. CDA의 활성이 낮은 사람은 solid tumor의 항종양제로부터 중증/치사 독성을 경험할 가능성이 높다.^{1 2} 그리고 일부 암의 경우, CDA 발현이 거의 없거나 없는 종양 환자가 더 나은 예후를 보인다.⁴ 따라서 환자의 genome과 eQTL을 비교하여 CDA의 발현량을 추측해볼 수 있을 것이며, CDA의 발현량을 통해 solid tumor의 항종양제에 대한 내성 정도와 일부 암에 걸릴 경우 그 예후를 미리 추측할 수 있다.

본문에서 중심으로 다루지는 않았지만 chr5의 CLTB 유전자 또한 발견된 eQTL과 association을 가졌다. chr5에 존재하는 rs116109937에 의해 CDA와 CLTB 두 유전자의 발현량이 유의미하게 변화한다는 결과가 나왔다(CLTB : p=1.723e-06). rs116109937의 chromatin interaction을 확인한 결과, chr5 내부에서 매우 넓게 상호작용이 발생함을 확인할 수 있었다. chia-pet으로 확인한 결과 GM12878 cell line에서 RAD21을 타겟으로 하는 chromatin interaction이 chr5:143583276-143583772 (hg19)부터 chr5:144617287-144617840까지 확인되었다. hichip으로 확인한 결과 GM12878에서 cohesin을 타겟으로 하는 상호작용이 chr5:143580000-143590000에서 chr5:144610000-144620000 (hg19)로 확인

되었다. 상호작용하는 해당 범위 안에 유전자가 많지는 않지만 어떤 작용에 의해 rs116109937가 chr5를 넓게 조절할 수 있는지에 대한 추가적인 연구가 필요할 것이다.

또한 CDA의 eQTL 중 chr2의 DTNB intron에 포함된 eQTL은 zinc finger의 가능성 또한 가진다. zinc finger의 타겟이 CDA일지 tool을 통해 알아보려 했지만 유의미한 결과가 나오지는 않았다. 하지만 CDA의 발현에 DTNB가 연관되어있을 수 있기 때문에 이에 대한 추가적인 연구가 필요할 것이다.

5. Conclusion

chr1의 rs12726436, rs12731069, rs12752560, rs61781229, rs35859441, rs61781231, rs6665553, rs818197, rs818196, rs577042, rs517483이 속한 eQTL은 TSS 또는 promoter 또는 enhancer로 기능할 수 있고, 그 타겟은 CDA라 추측된다.

특히, rs12726436은 CDA의 TSS인 동시에 promoter로 기능할 것으로 예상되기 때문에, minor allele(A)의 수가 증가할수록 CDA 발현량이 증가함으로 추측된다. minor allele일 때 MYBL1::MAX가 CDA의 promoter에 더 강하게 bind할 수 있으므로, MYBL1::MAX가 CDA의 발현을 upregulation시켜 minor allele의 수가 증가할수록 CDA 발현량이 증가하는 것일 것이다.

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.