Disentangler A Visualization Technique for Linkage Disequilibrium Mapping Using Multi-allelic Loci

Kumasaka N[1] Okada Y[1] Takahashi A[1] Kubo M[1] Nakamura Y[2] Kamatani N[1]

[1] Center for Genomic Medicine RIKEN [2] Institute of Medical Science University of Tokyo

Introduction

Linkage disequilibrium (LD) mapping using multi-allelic genetic loci (such as *HLA* alleles) is occasionally required to identify susceptibility haplotypes for complex diseases [1]. Here we propose a new visualization technique to map LD with multi-allelic loci. A statistical method to infer underlying haplotypes between these loci is developed and the multiple correspondence analysis (MCA, also known as homogeneity analysis) is further applied, which provides an optimal default of the inferred haplotypes on a display. The method is also general and applicable for single-nucleotide polymorphisms and allele-specific copy number polymorphisms, or mixture of those in a single display.

Results

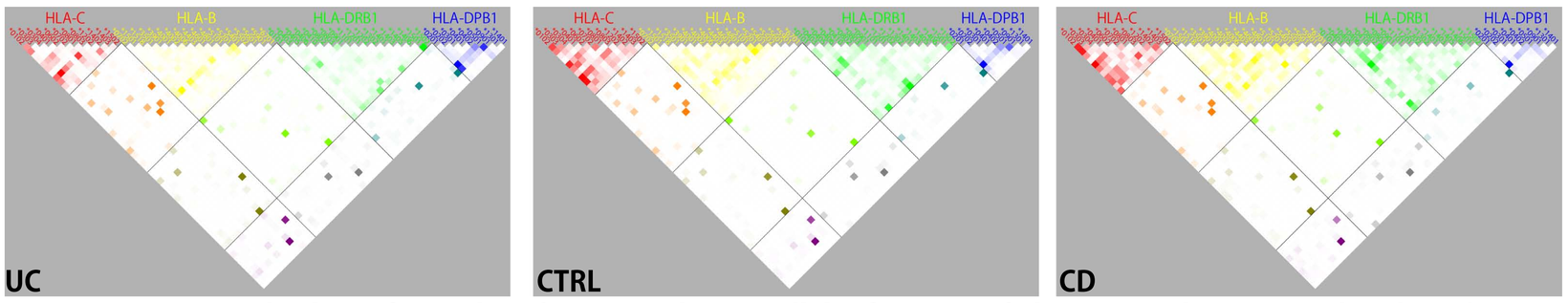


Figure 1. Heatmap displays of LD index, r^2, among the HLA alleles for cases of ulcerative colitis (UC) and Crohn's disease (CD) along with controls (CTRL). The block with different color divided by the black solid lines indicates each of the 4 HLA genes; HLA-C (red), HLA-B (yellow), HLA-DRB1 (green) and HLA-DPB1 (blue). The darker the color shading is the stronger the strength of LD is.

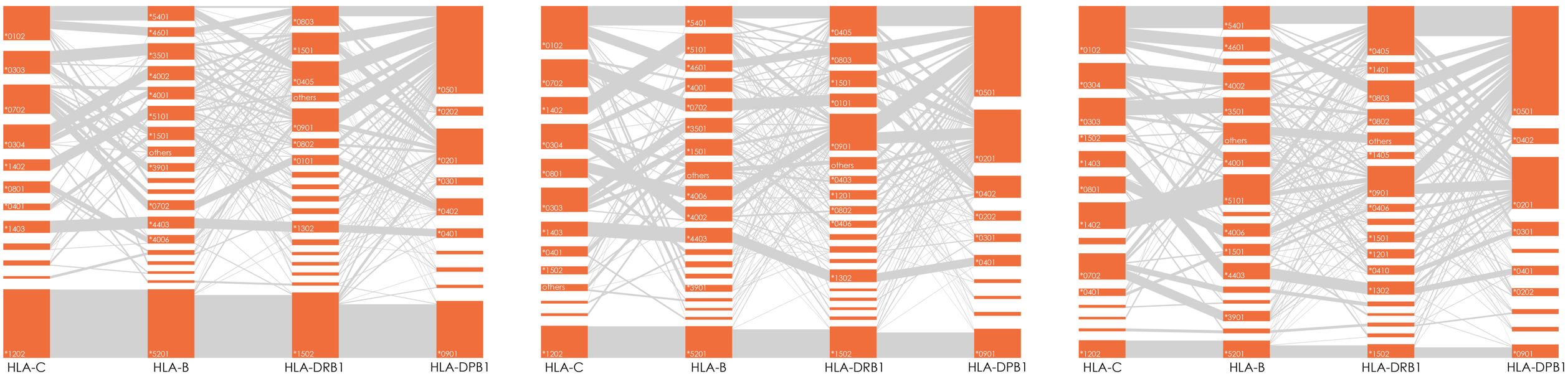


Figure 2. Haplotype structure of HLA alleles for cases of ulcerative colitis (UC) and Crohn's disease (CD) along with controls (CTRL). The vertical stacked bar indicates each of the 4 HLA genes,, and the queues of the bars correspond to their physical order in the MHC region. A tile of a bar indicates an HLA allele, and a segment connects 2 alleles on adjacent genes. The height of the tile and the thickness of the segment correspond to the allele frequency of the HLA allele and haplotype frequency between the 2 HLA alleles, respectively. The existence of the common haplotype consisting of Cw*1202, B*5201 and DRB1*1502 is clearly shown, whose frequency decays from in order of UC cases, the controls, and CD cases, representing its opposite directions of effects on UC and CD (see [1] for more details).

Table. Associations of the haplotype consisting of HLA alleles associated between UC cases and CD cases (Table 4 in [1])

Haplotype			Frequency[a]			UC vs CD		UC vs control		CD vs control	
			UC (n = 372)	CD (n = 372)	Control (n = 905)	OR (95% CI)[b]	*P* value[b]	OR (95% CI)[b]	*P* value[b]	OR (95% CI)[b]	*P* value[b]
Haplotype for Cw∗1202, B∗5201, DRB1∗1502											
Cw∗1202	B∗5201	DRB1∗1502	0.27	0.054	0.12	6.58 (4.60 − 9.42)	1.1×10^{-33}	2.65 (2.14 − 3.29)	4.0×10^{-21}	0.40 (0.28 − 0.57)	1.1×10^{-7}
Cw∗1202	B∗5201	—	0.025	0.022	0.013	1.51 (0.76 − 2.97)	.15	2.24 (1.21 − 4.14)	.0080	1.49 (0.79 − 2.81)	.21
—	—	DRB1∗1502	0.011	0.0056	0.012	2.64 (0.81 − 8.65)	.030	1.10 (0.49 − 2.46)	.54	0.42 (0.15 − 1.19)	.067
—	—	—	0.69	0.92	0.85						
Haplotype for Cw∗1402, B∗5101											
Cw∗1402	B∗5101		0.050	0.11	0.072	0.42 (0.28 − 9.42)	1.8×10^{-5}	0.67 (0.46 − 0.98)	.039	1.62 (1.21 − 2.17)	.0010
—	B∗5101		0.013	0.016	0.017	0.78 (0.33 − 1.81)	.57	0.78 (0.38 − 1.61)	.48	1.01 (0.51 − 1.98)	.90
—	—		0.94	0.87	0.91						

NOTE. HLA alleles other than Cw ∗1202, B∗5201, DRB1∗1502, or Cw∗1402, B∗5101 are pooled and denoted as "—."
n, Number of subjects enrolled in the analysis.
[a]Haplotype with >0.5% of frequency in the controls are indicated.
[b]Obtained by the comparison of haplotype frequencies between each of the haplotype and the haplotype with the highest frequency.

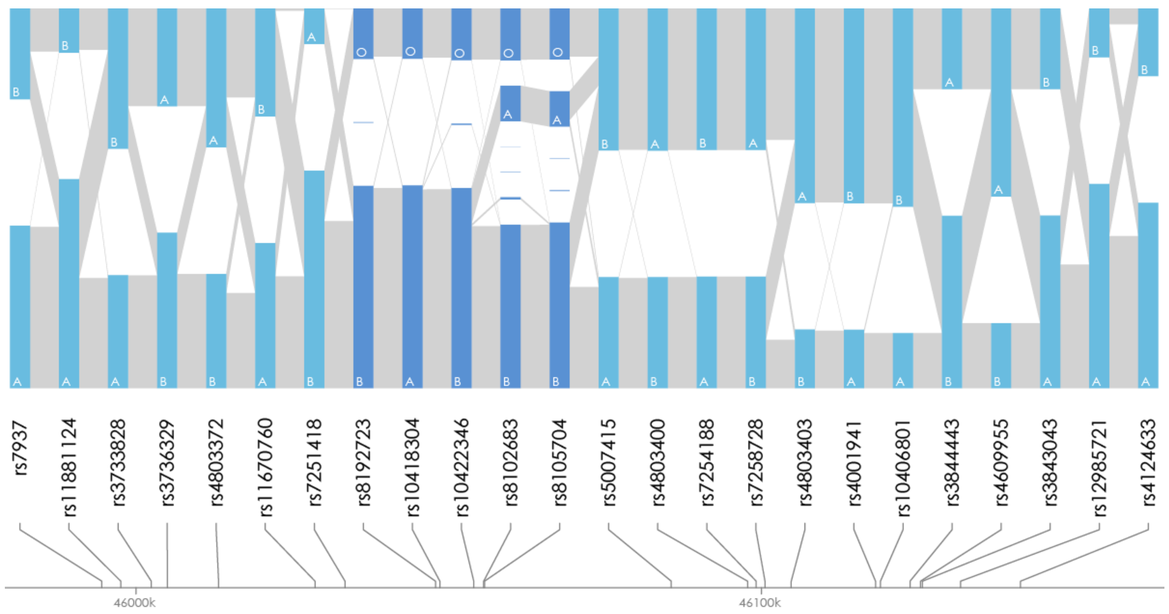


Figure 3. Haplotype structure among SNPs (light blue) and allele-specific CNPs (navy) on chromosome 19 Alleles of each SNP locus are indicated by "A" and "B", and the common deletion at CNP locus is indicated by "O". Note that allele-specific CNPs were genotyped by using PlatinumCNV software [3].

Method Horizontalization Criterion

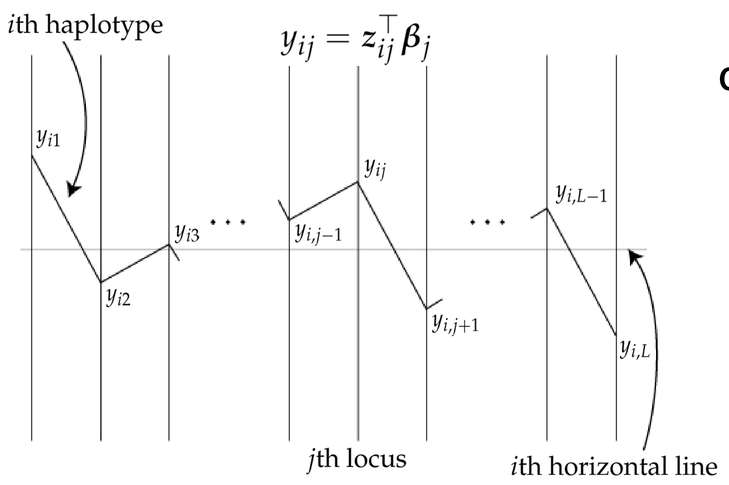
z_{ij} : indicator vector of haplotype i at locus j
β_j : scale parameters of alleles at locus j
ξ_i : ith horizontal line
π_i : ith haplotype frequency
$p_l^{(j)}$: lth allele frequency at locus j
$p_{lm}^{(jk)}$: haplotype frequency of lth allele at locus j and mth allele at locus k
A_j : number of possible alleles at locus j

$$A = \begin{pmatrix} S_{11} & \cdots & S_{1L} \\ \vdots & \ddots & \vdots \\ S_{L1} & \cdots & S_{LL} \end{pmatrix},$$

$$B = \begin{pmatrix} S_{11} & & 0 \\ & \ddots & \\ 0 & & S_{LL} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_L \end{pmatrix}$$

$$\text{and } S_{jk} = \begin{pmatrix} s_{11}^{(jk)} & \cdots & s_{1A_k}^{(jk)} \\ \vdots & \ddots & \vdots \\ s_{A_j1}^{(jk)} & \cdots & s_{A_jA_k}^{(jk)} \end{pmatrix}$$

where $s_{lm}^{(jk)} = p_{lm}^{(jk)} - p_l^{(j)} p_m^{(k)}$



Optimization Problem

$$\begin{cases} \sum_{ij} \pi_i (y_{ij} - \xi_i)^2 \underset{\{\xi_i\},\{\beta_j\}}{\rightarrow} \min \\ \text{s.t.} \sum_{ij} (y_{ij} - \bar{y}_j)^2 = \text{const} \end{cases}$$

$$\underset{[2]}{\Leftrightarrow} \begin{cases} A\hat{\boldsymbol{\beta}} = \lambda B \hat{\boldsymbol{\beta}} \\ \hat{\xi}_i = c \sum_{j=1}^{L} z_{ij}^{\top} \hat{\beta}_j \end{cases}$$

Web site

http://kumasakanatsuhiko.jp/projects/disentangler/

References

1. Okada Y et.al. (2011) HLA-Cw*1202-B*5201-DRB1*1502 haplotype Increases risk for ulcerative colitis but reduces risk for Crohn's disease, *Gastroenterology* **141**(3) 864-71.
2. Kumasaka N et.al. (2008) High-dimensional data visualisation: the Textile Plot, *Computational Statistics & Data Analysis* **52** 3616-3644.
3. Kumasaka N et.al. (2011) PlatinumCNV: a Bayesian Gaussian mixture model for genotyping copy number polymorphisms using SNP array signal intensity data, *Genet. Epidemiol.* In press.

Acknowledgement

We are grateful to the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for supporting our study.
We also thank all the staffs of Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN for their contribution to SNP genotyping.

Disclosure

This work has not been published.
Authors certify that there is no conflict of interest with any financial organization regarding the material discussed in this poster.
This work was supported by the Ministry of Education, Culture, Sports, Sciences and Technology of the Japanese government.
All individuals enrolled in the study gave their written informed consent, and approval was obtained from the ethical committees at Kyushu University, Social Insurance Chuo General Hospital, and RIKEN Yokohama Institute.