

# エントピー正規化と高次元データ可視化技術によるHLA遺伝子多型構造の評価

岡田 随象 (ゆきのり)

URL: <http://plaza.umin.ac.jp/~yokada/>  
E-mail: yokada.brc at tmd.ac.jp

東京医科歯科大学 大学院医歯学総合研究科 疾患多様性遺伝学分野

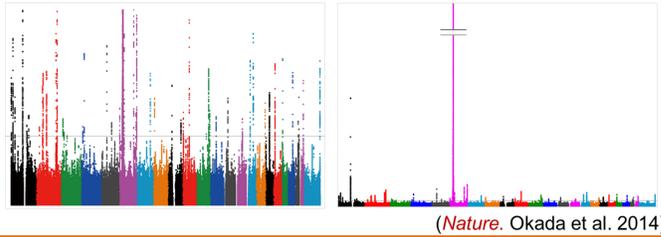
理化学研究所 統合生命医学研究センター 統計解析研究チーム

ヒト主要組織適合性複合体(MHC)領域内に存在するHLA遺伝子は多彩なヒト疾患リスクを有するが、複雑な領域内遺伝子多型構造がリスク遺伝子多型同定の障壁となっていた。HLA imputation法は、SNPアレイデータに基づきHLA遺伝子データをコンピューター上で高精度に推定する遺伝統計解析手法であり、既存のゲノムワイド関連解析に対する網羅的HLA遺伝子解析を追加費用なしで実施可能にする。我々は、HLA imputation法実施時に必要な参照HLAデータを日本人健康者908名を対象に作成し、同手法の日本人集団への実装を行った。また、ハプロタイプ頻度分布に対する情報量エントピーの正規化指標 $\epsilon$ (イプシロン)や高次元データ可視化手法Disentanglerを活用してHLA遺伝子間の連鎖不平衡関係を検討し、MHC領域全域にわたる人種特異的なHLAハプロタイプの存在を明らかにした。得られた参照データを用いて、甲状腺機能の異常亢進を伴う自己免疫疾患であるバセドウ病の日本人集団ゲノムワイド関連解析( $n = 9,003$ )にHLA imputation法を適用した。結果、複数のクラスIおよびクラスII HLA遺伝子(HLA-DPB1、HLA-A、HLA-B、HLA-DRB1)の機能的部位に位置するアミノ酸多型が独立にバセドウ病発症リスクを有することが判明した。特に、HLA-DPB1遺伝子の35番目のアミノ酸残基が最も強いリスクを有していた(オッズ比=1.42)。今回取得した日本人集団HLA参照データは、他疾患へのHLA imputation法の適用による疾患バイオマーカーの同定や、次世代シーケンサーを用いたHLAアレル配列決定時の参照データとしての活用が期待される(Okada Y et al. *Nature Genetics* 2015)。

## ~ MHC領域におけるヒト疾患罹患リスクとHLA imputation法によるfine-mapping ~

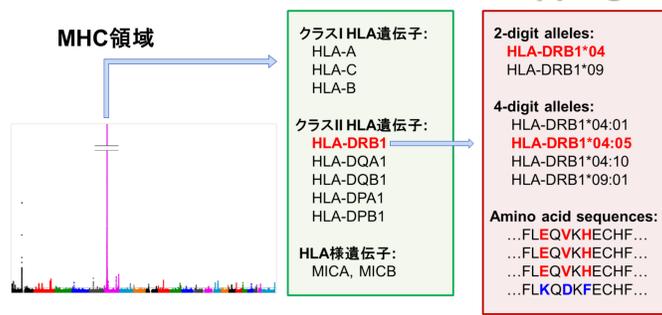
### MHC領域における疾患罹患リスク

関節リウマチ (RA) におけるゲノムワイド関連解析結果  
(19,234 RA cases vs 61,565 controls)



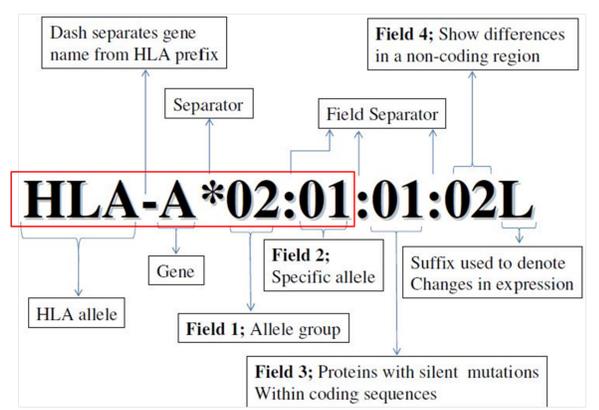
- 6番染色体短腕上のMHC領域は、自己免疫疾患、悪性腫瘍、感染症、精神疾患といった多彩なヒト疾患に対するリスクとの強い関連を有する。
- 関節リウマチ(Rheumatoid arthritis; RA)では、遺伝子数の1/3を説明。
- MHC領域内は構造が複雑であり、複数のHLA遺伝子やHLA様遺伝子が存在するため、感受性遺伝子の同定(fine-mapping)が困難であった。

### MHC領域におけるリスクfine-mapping



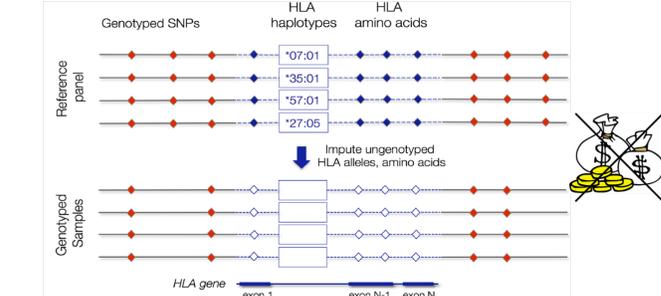
- MHC領域内には、複数のHLA遺伝子やHLA様遺伝子が存在する。
- 各HLA遺伝子が複数のHLAアレル(2-digit/4-digit classical alleles)やHLAアミノ酸多型を持つ。
- 複数遺伝子、アレル、アミノ酸多型を網羅的に扱う遺伝統計解析が必要。

### HLAアレルの命名方法



- 4桁の古典的HLAアレルが、特定のHLAアミノ酸配列に対応

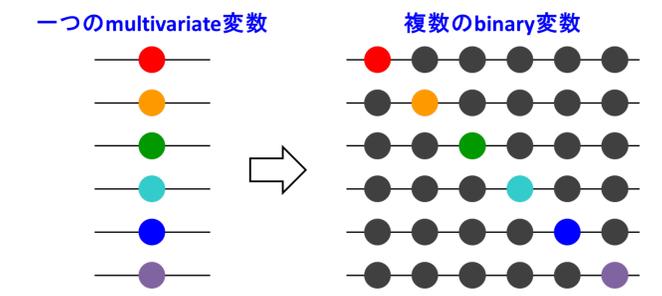
### HLA imputation法によるHLA遺伝子多型推定



- HLA遺伝子のタイピングは高額(=~10万円/サンプル)。
- HLA imputation法を用いることにより、通常のSNPジェノタイプデータから「追加費用なし」で、HLA遺伝子多型をコンピューター上で高精度で推定することが可能になる。

(PLoS One Jia et al. 2013)

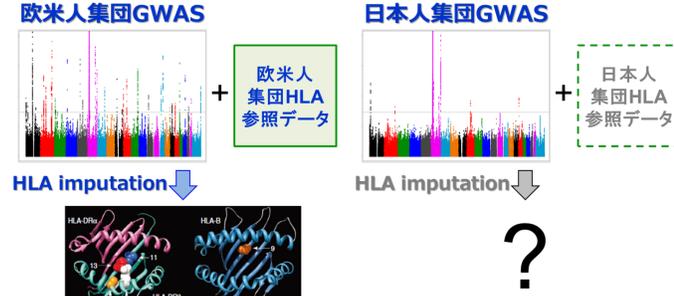
### HLA imputation法における遺伝子多型の扱い



- Imputation作業そのものは既存のソフトウェアを使用(BEAGLE)。
- マルチアレルなHLA遺伝子多型を「一つのmultivariate変数」ではなく「複数のbinary変数」として扱った結果、従来以上の性能と、HLAアミノ酸多型imputationへの拡張が可能になった。

(PLoS One Jia et al. 2013)

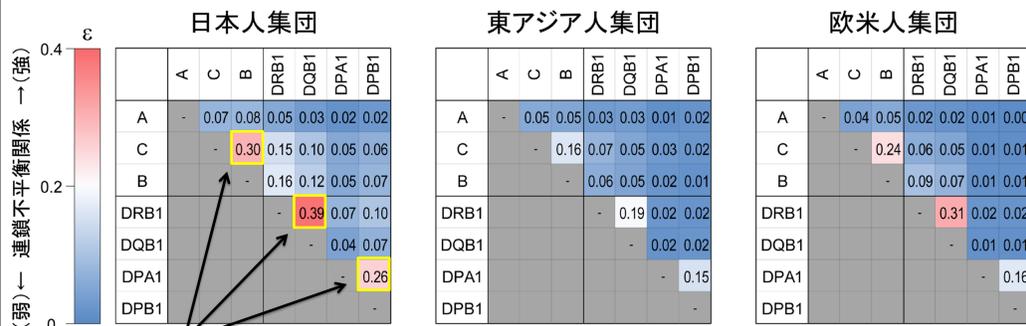
### 日本人集団におけるHLA imputation法の実装



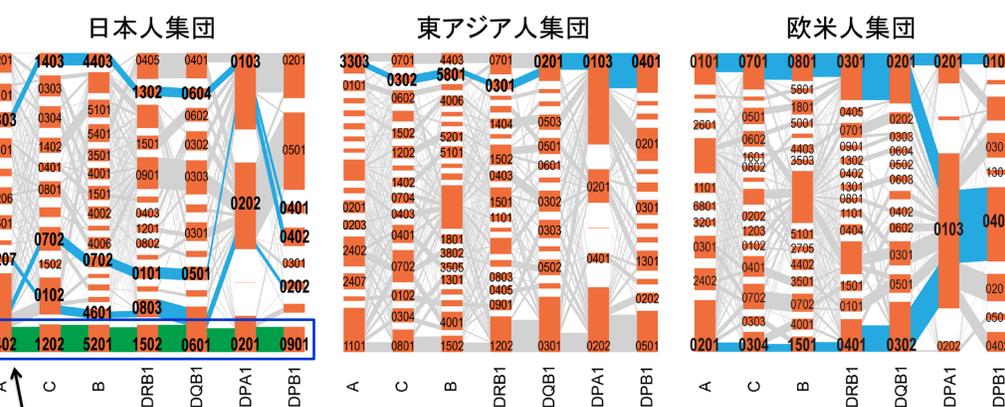
- HLA imputation法の実施には、参照データ(reference panel)が必要。
- 日本人集団においてHLA imputation法を実装するため、日本人集団を対象とした参照データを新たに作成( $n = 908$ )。
- 高精度なHLA imputationを実施可能に(4-digitアレル一致率 $\geq 92\%$ )。

## ~ ビッグデータ可視化手法とHLA遺伝子構造 ~

### 情報量エントピー正規化指標 $\epsilon$ (イプシロン)によるHLA遺伝子間の連鎖不平衡(LD)関係の評価



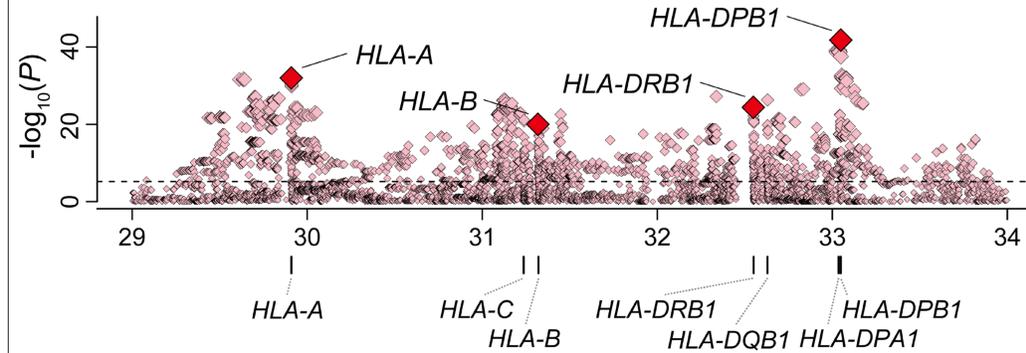
強固なLDが日本人で顕著 高次元データ圧縮技術DisentanglerによるHLA遺伝子ハプロタイプ構造の可視化



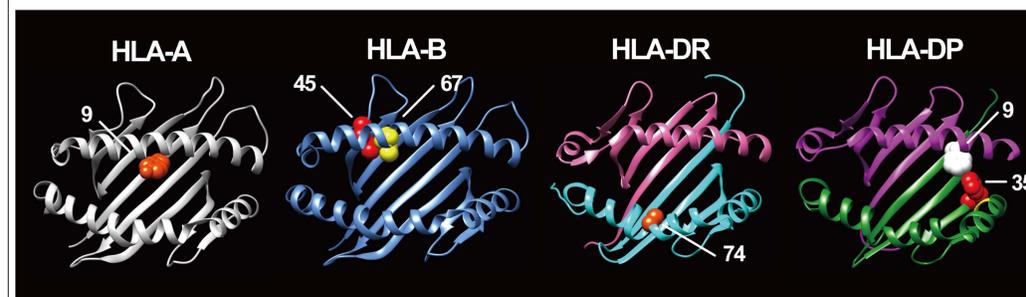
日本人集団に特異的なHLA遺伝子ハプロタイプの発見!!  
(HLA-A\*24:02-HLA-C\*12:02-HLA-B\*52:01-HLA-DRB1\*15:02-HLA-DQB1\*06:01-HLA-DPA1\*02:01-HLA-DPB1\*09:01)

## ~ バセドウ病リスクHLA遺伝子多型の同定 ~

### 日本人集団のバセドウ病ゲノムワイド関連解析におけるMHC領域内の疾患罹患リスク ( $n = 9,003$ )



6番染色体上の位置 (Mbp)  
バセドウ病の疾患罹患リスクを有するHLA遺伝子アミノ酸変異



日本人集団HLA参照データは、科学技術振興機構(JST)バイオサイエンスデータベースセンター(NBDC)を通じて一般公開中  
<http://humandbs.biosciencedbc.jp/hum0028-v1>

