

# 日本人集団においてアレル頻度の経験分布は染色体ごとに異なる

Empirical distributions of allele frequencies in Japanese population are different for each chromosome.

川口 章夫<sup>1</sup>、中本 和典<sup>2</sup> 1. 山梨大学医学部附属病院 山梨県地域医療支援センター 2. 山梨大学 医学部 医学教育センター

## 1. 背景

アレル頻度の経験分布が人種により異なることは複数報告されている<sup>1), 2)</sup>。独立なSNVの数は人種ごとに異なっている(低頻度アレル頻度の分布、LDの大きさなど)。それらに基づき、GWASにおける人種ごとに適したゲノムワイドの有意水準も提案されている<sup>3)</sup>。

しかし、染色体ごとのアレル頻度の経験分布の違いは検討されていない。そこでゲノムの多様性を「常染色体ごとのアレル頻度の経験分布」から検討する。

## 2. 方法

東北メディカル・メガバンク機構(ToMMo)の健常な日本人の「日本人ヒト全ゲノム解析に基づく高精度の住民ゲノム参照パネル(2,049人)」(2KJPN)<sup>4)</sup>のアレル頻度情報を利用した。

2KJPNは、約2,800万のSNVを有し、それらの位置・アレル頻度・アレル数の情報が含まれている。

ToMMoからアレル頻度情報をダウンロード(<https://ijgvd.megabank.tohoku.ac.jp/download/>)し、LinuxコマンドとRを使いデータ処理を行った。

“TSVファイルの構造”

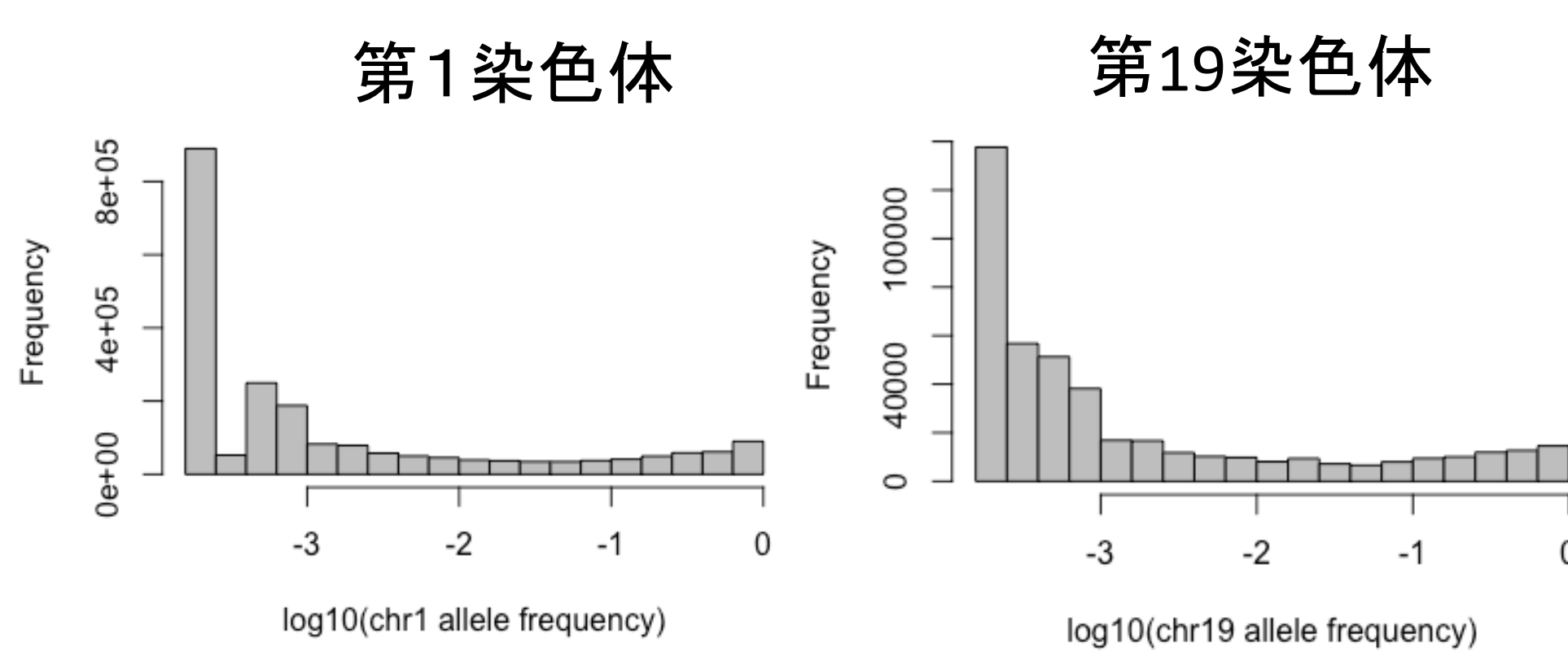
```
1 ~ 64行目 : コメント(#から始まる)
65行目      : 空行
66行目      : 列名
67行目 ~    : 観測値(アレル頻度など)
```

日本人においては多様性のないSNVも含まれているため、それらのSNVは除去した。

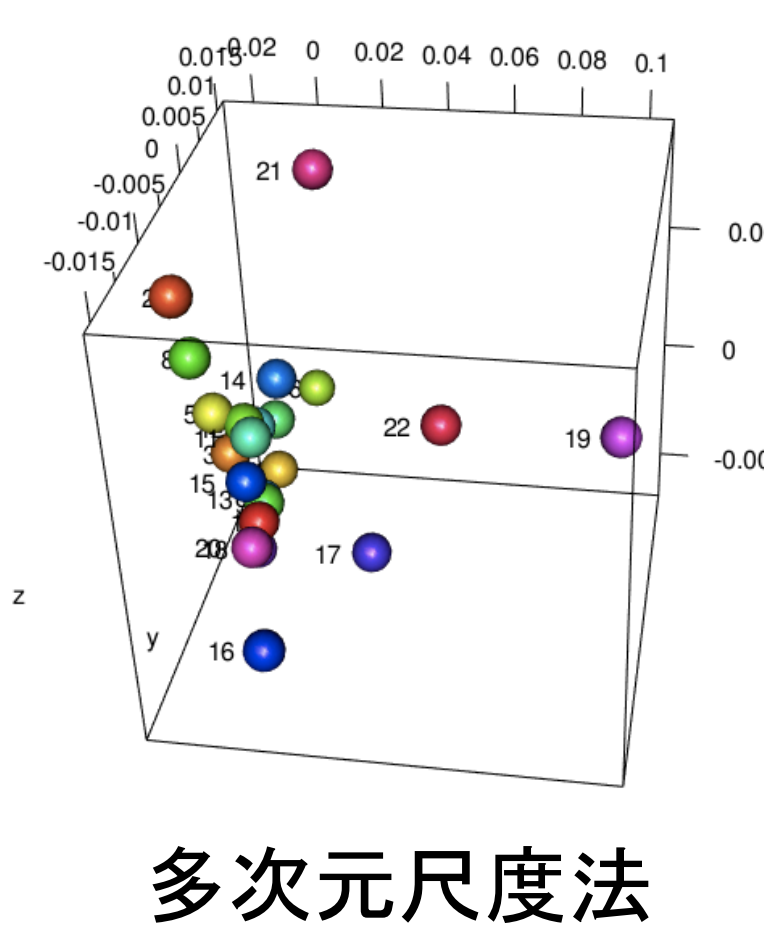
列名	Chromosome	Position_hg19	rsSNP#	Cytoband
	Variation_type	Ref_allele	Ref_allele_freq	
	Ref_allele_count	Alt_allele	Alt_allele_freq	
	Alt_allele_count	Total_allele_count		
	#Samples	Gene_annotation		

## 3. 結果

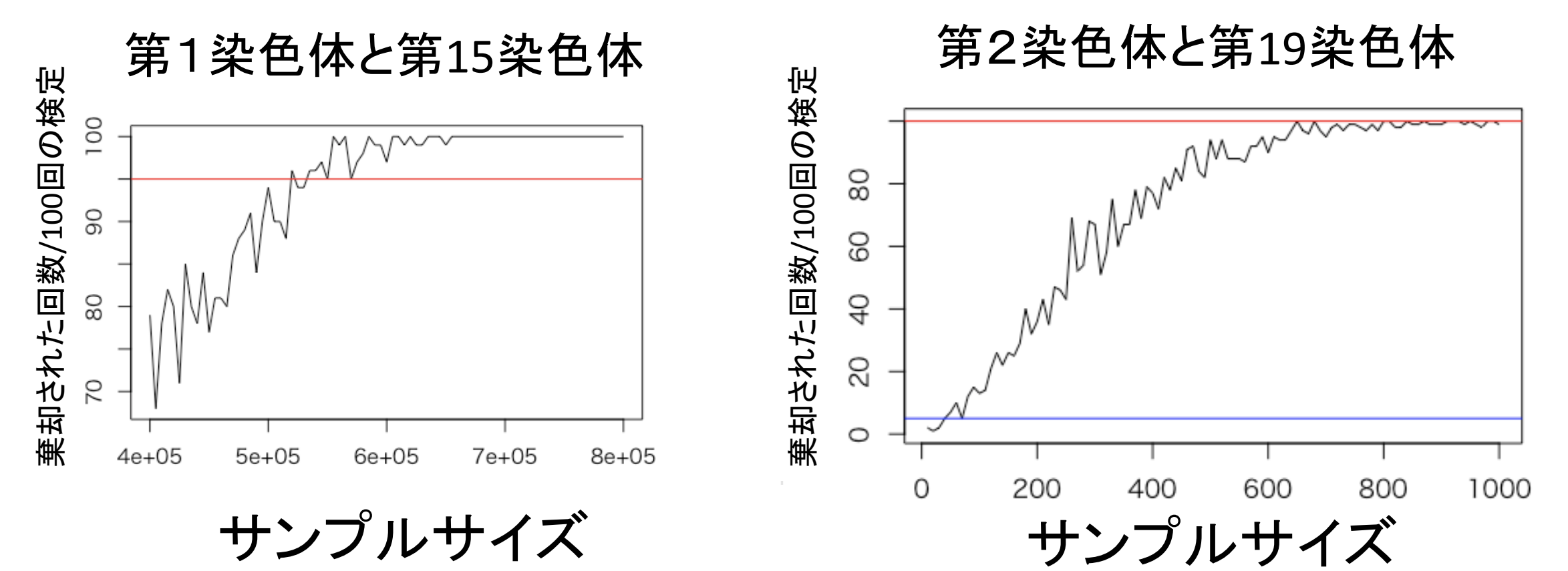
2KJPNのアレル頻度情報より、代表的なアレル頻度の経験分布に違いのあるアレル頻度の経験分布のヒストグラムを示した。



染色体間の距離としてKS検定量を用い、多次元尺度法で染色体間のアレル頻度の親近性を示した。



さらに、全ての染色体から抽出する標本のサンプルサイズを変えながらKS検定を行い棄却されるまでのサンプルサイズの違いから染色体間のアレル頻度の親近性を検討した。

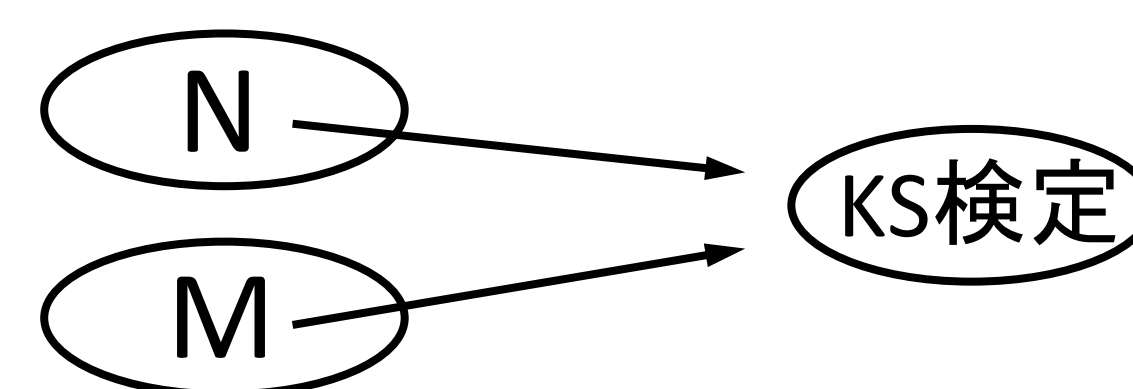


1) 全ての染色体の組み合わせでコルモゴロフ・スミルノフ検定(KS検定)が棄却された。つまり、KS検定においてアレル頻度の経験分布は染色体ごとに異なる。

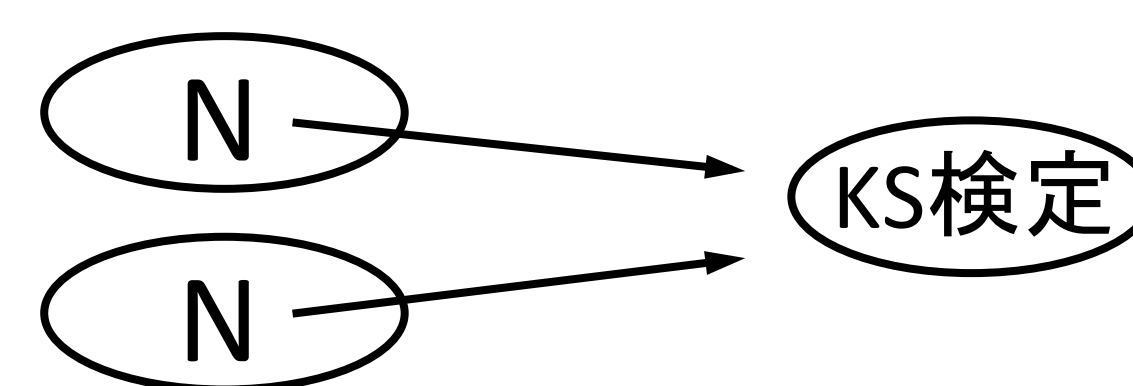
2) また、同一染色体同士でKS検定を行っても棄却されないことも確認した。

3)、4) 異なる、または、同一の染色体からサンプルサイズを変えながらKS検定を行った。

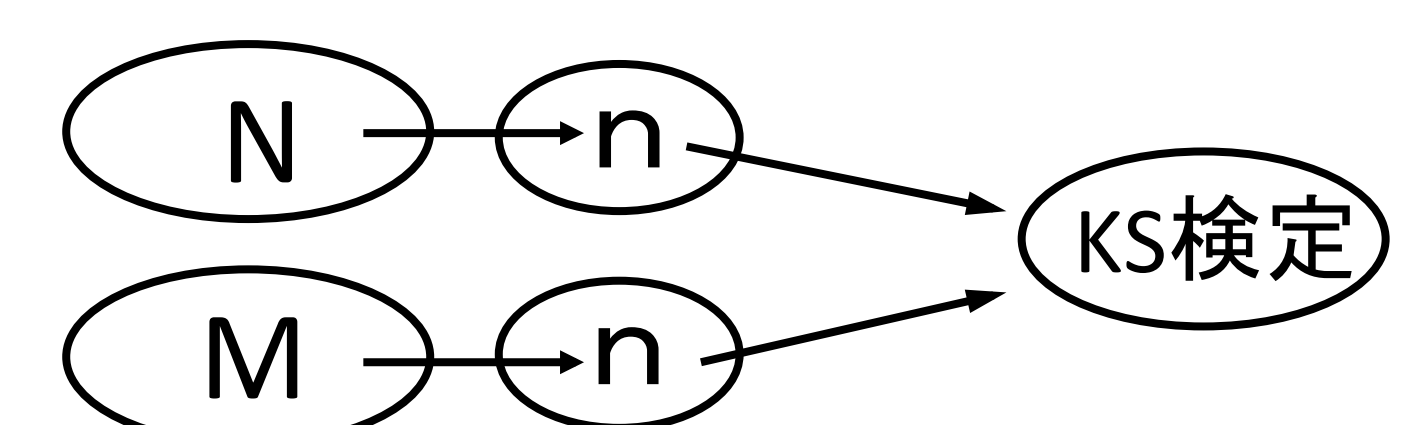
1) 全ての染色体の組み合わせでのKS検定は、“棄却された”(N, Mは異なる染色体全体のアレル頻度)



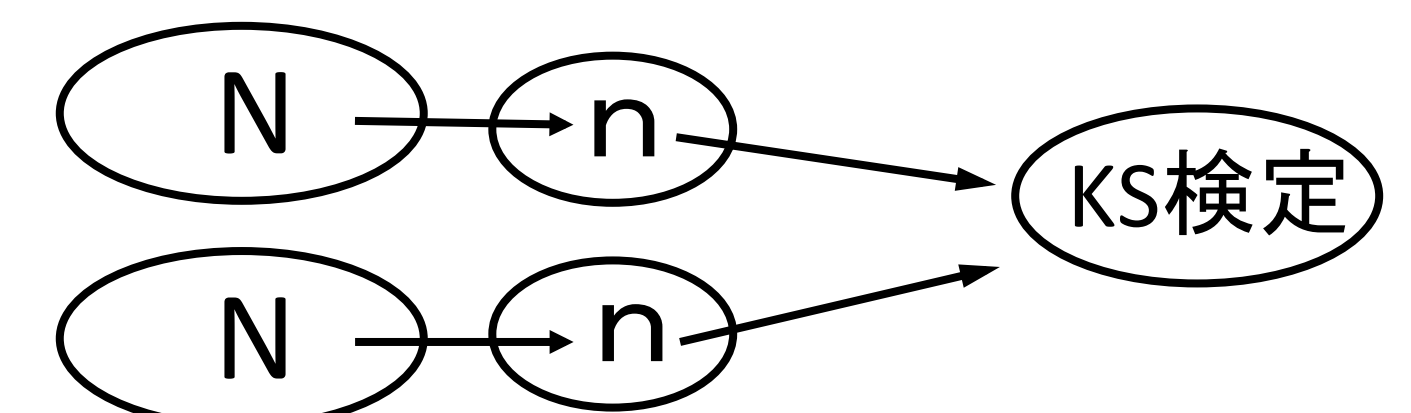
2) 同一染色体同士でのKS検定は、“当然ながら棄却されなかった”



3) 異なる染色体からサンプルサイズを変えながらKS検定は、“親近性により棄却されるサンプルサイズが異なった”



4) 同一染色体からサンプルサイズを変えながらのKS検定は、“サンプルサイズを大きくしても棄却されなかった”



## 4. 結論

我々は最新(2017年7月時点)かつ人数の多い日本人リファレンスパネル(2KJPN)を用い、KS検定において常染色体ごとにアレル頻度の経験分布が異なることを明らかにした。

## 5. 考察

- 2KJPN以外に1KJPN・3.5KJPN、1000 Genomes Projectのdata setにおいてもアレル頻度の経験分布が染色体ごとに異なるかを確認する必要がある。
- 染色体ごとのアレル頻度の分布関数の推定を行っている。
- 今回の結論から、GWASの閾値を染色体ごとに決めることで、より適切なゲノムワイドの有意水準を提案することも今後の課題である。

## 引用文献

- Li et al. *Nat. Genet.* 2010  
Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants.
- Gudbjartsson et al. *Nat. Genet.* 2015  
Large-scale whole-genome sequencing of the Icelandic population.
- Kanai et al. *Journal of Human Genetics* 2016  
Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set.
- Nagasaki et al. *Nature Communications* 2015  
Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals.

ご質問・ご意見

[akiok@yamanashi.ac.jp](mailto:akiok@yamanashi.ac.jp)

日本人類遺伝学会第62回大会

利益相反状態の開示  
開示すべき利益相反状態はありません。