

# オープンデータからCOVID-19の 遺伝的死亡リスクを探索する 統計モデルの開発

第70回日本感染症学会東日本地方会学術集会 優秀演題選出

上智大学大学院理工学研究科

修士2年 佐藤 茉莉香

北海道札幌市出身

東京女子大学 数理科学科 情報理学専攻を卒業

2020.03  
東京女子大学現代教養  
学部 数理科学科卒  
複雑系科学  
卒論タイトル：  
『炎上の数理モデル』

2020.04  
上智大学大学院  
理工学研究科入学  
統計学を学ぶ  
・金融工学  
(~2020.07)

2020.08~  
・**遺伝統計学**を学ぶ  
ゲノムワイド関連研究  
疾患や薬の副作用などの先天的な  
リスクをゲノム情報から探索する  
遺伝統計学的手法

修士の立場で臨床データを取得することは  
難しい。オープンデータベースであれば  
誰にでも取得可能であり、今後同じような  
ウイルスが蔓延した際に誰でも研究可能  
である。

SNS  
Twitter

学士

修士

COVID-19について  
修士という立場で  
何か研究できないか？

インターネット上に公開  
されたデータ（オープン  
データベース）を使って  
新型コロナウイルスの  
先天的な要因による重症化  
リスクを見つけることが  
できないか

2019.12  
新型コロナウイルス  
発生

2020.02  
COVID-19  
Pandemic

欧米の重症化率あるいは死亡率が  
アジアなどに比べて高いとの報道が  
多く見られる

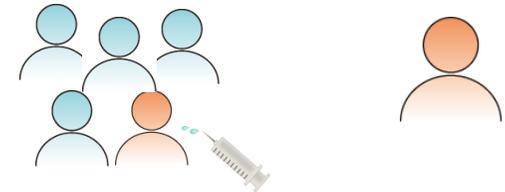
環境的要因

遺伝的要因

予めリスクタイプが  
わかれば…

✓ リスク保因者は優先的に  
ワクチンを接種することで  
重症化率をへらすことができる

✓ 3密からの  
回避行動への  
意識付けができる



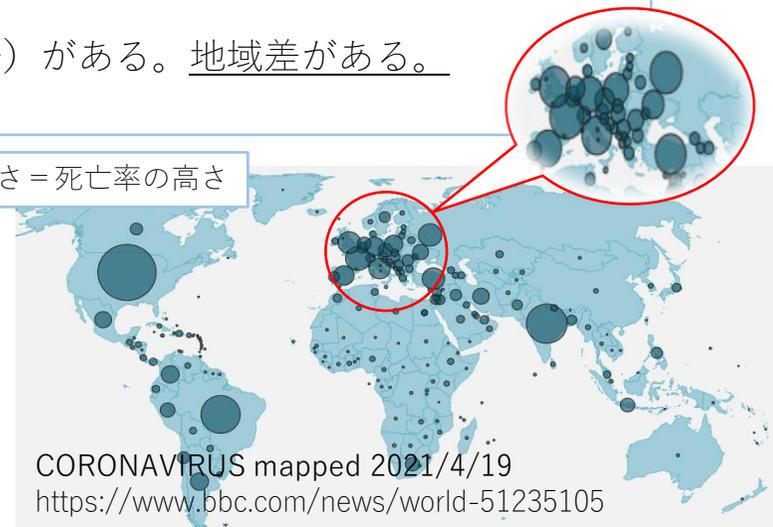
# 概要

患者データを使わずにオープンデータベースのみで  
新型コロナウイルス感染による死亡リスクとしてHLA遺伝子の関与を検討する

HLA (Human Leukocyte Antigen = ヒト白血球抗原)

白血球の血液型と呼ばれ、16,000以上の種類 (アレル) がある。地域差がある。  
両親から1対ずつ受け継ぐ。

円の大きさ = 死亡率の高さ



最も重大な  
有害事象

表現型が明確

なぜ死亡リスクか

各国の死亡者数は  
正しい情報が  
公開されている

診断基準が明確

死亡率は近隣諸国でも  
異なるように見える  
(例：ヨーロッパ)

なぜHLA遺伝子か

遺伝的要因を仮定するのであれば、  
地理的に近いが遺伝的多様性が  
ある領域がターゲット

盛んに研究され  
データベースも豊富

使用したオープンデータベース

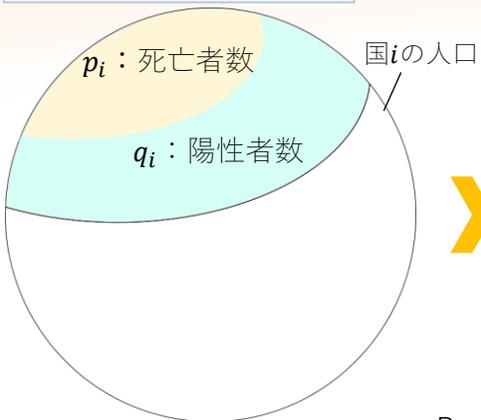
- WHO (<https://covid19.who.int/>)  
2019年12月30日～2021年4月21日までのデータ  
計26ヶ国
- Allele frequency net database  
(<http://www.allelefrequencies.net/hla6006a.asp>)  
1,000名以上が登録されたデータセット (110セット)  
対象とするHLA領域はHLA-A, C, B, DPB1, DQB1, DRB1

HLAアレルと病気は密接に関係しており、我々の研究では新型コロナウイルス感染による  
重症化 (死亡) リスクに関連するHLAを同定する遺伝統計学的数理モデルを構築した。

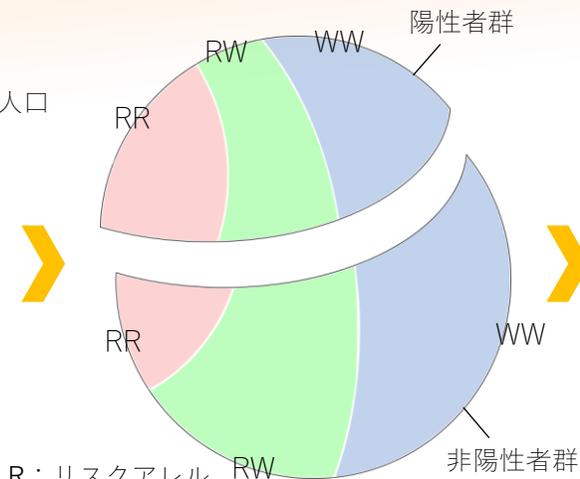


# 浸透率推定のための遺伝統計学的数理モデル

## 遺伝統計学的思考



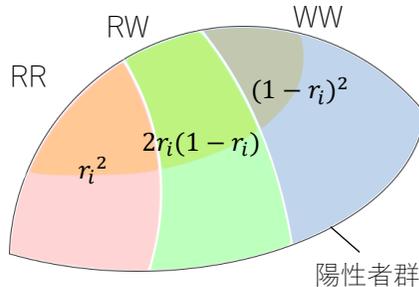
$p_i, q_i$  はWHOのデータを使用



R : リスクアレル  
W : その他のアレル

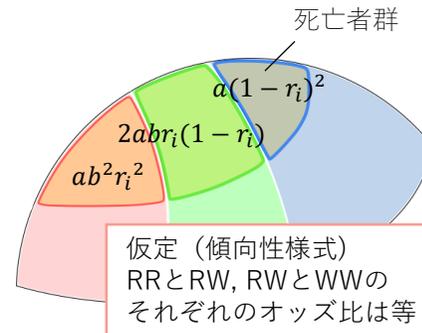
仮定 : 陽性者群、非陽性者群でそれぞれ  
ハーディーワインバーク平衡が成立する

仮定 : コロナウィルスはHLAアレル  
に対してランダムに感染する



ハーディーワインバーク平衡より、  
遺伝型の頻度が表せる。  
 $r_i$  : 対象アレルの国*i*におけるアレル頻度

a : WWの浸透率  
b : Rが一つ増えるごとの死亡オッズ  
b>1ならリスク、b<1なら非リスクアレル



仮定 (傾向性様式)  
RRとRW, RWとWWの  
それぞれのオッズ比は等しい

・浸透率

RR	RW	WW
$ab^2$	$ab$	$a$

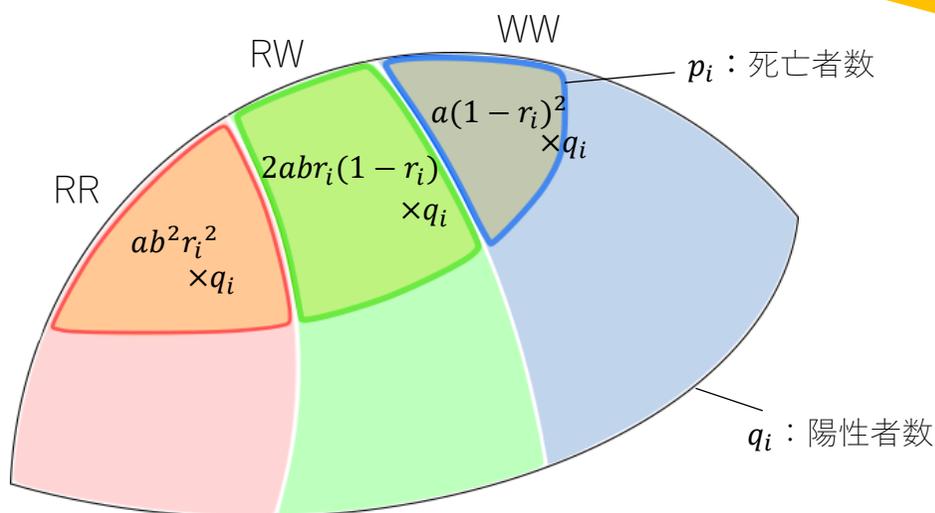
$\times b$        $\times b$

浸透率とは  
各遺伝型における  
死亡確率

a : WWの浸透率  
b : Rが一つ増えるごとの死亡オッズ  
 $r_i$  : 対象アレルの国*i*におけるアレル頻度 (アレルの割合)

$$ab^2r_i^2 + 2abri(1-r_i) + a(1-r_i)^2 = \frac{p_i}{q_i} \quad (\text{死亡率})$$

$$\Leftrightarrow a\{(b-1)r_i + 1\}^2 = \frac{p_i}{q_i} \quad \dots (\text{Eq1})$$



# 結果

HLAアレル (R)	WWの浸透率(a)		Rが一つ増えるごとの死亡オッズ(b)		各遺伝型の浸透率		
	Estimate (SE)	P	Estimate (SE)	P	WW	RW	RR
B*13:02	0.013 (0.003)	3.744E-04	8.373 (0.411)	1.150E-04	0.013	0.109	0.911
A*02:07	0.018 (0.003)	1.181E-04	6.478 (0.277)	1.468E-04	0.018	0.117	0.755
A*11:01	0.012 (0.003)	4.730E-04	4.647 (0.305)	1.831E-04	0.012	0.056	0.259
DQB1*05:02	0.015 (0.003)	4.261E-04	6.596 (0.333)	2.061E-04	0.015	0.099	0.653
B*46:01	0.016 (0.003)	6.739E-04	6.371 (0.269)	2.337E-04	0.016	0.102	0.649
C*01:02	0.016 (0.003)	6.644E-05	3.539 (0.265)	5.804E-04	0.016	0.057	0.200
C*08:01	0.019 (0.002)	8.681E-06	5.156 (0.353)	7.100E-04	0.019	0.098	0.505
DRB1*09:01	0.017 (0.002)	1.392E-06	3.273 (0.286)	9.786E-04	0.017	0.056	0.182
B*58:01	0.017 (0.003)	1.012E-05	5.669 (0.504)	3.623E-03	0.017	0.096	0.546

Point 2

例) B\*13:02の場合

国	アレル頻度	陽性者数	死亡者数
China	0.0775	103,165	4,855
Colombia	0.0092	2,552,937	66,156
Croatia	0.0376	297,973	6,399
Czechia	0.0475	1,590,124	28,124
France	0.0198	5,026,645	98,850
Germany	0.0344	3,044,016	79,088
India	0.0133	13,873,825	172,085
Ireland	0.0170	241,684	4,803
Israel	0.0448	836,469	6,312
Japan	0.0029	512,169	9,469
Netherlands	0.0250	1,364,025	16,822
Poland	0.0545	2,621,116	59,930
Republic of Korea	0.0309	111,419	1,782
Russian Federation	0.0492	4,666,209	104,000
Spain	0.0152	3,336,637	76,179
Switzerland	0.0244	622,510	9776
United States of America	0.0183	30,949,496	557,415

Point 1



Eq1へ代入して解a, bを求める

## Point 1

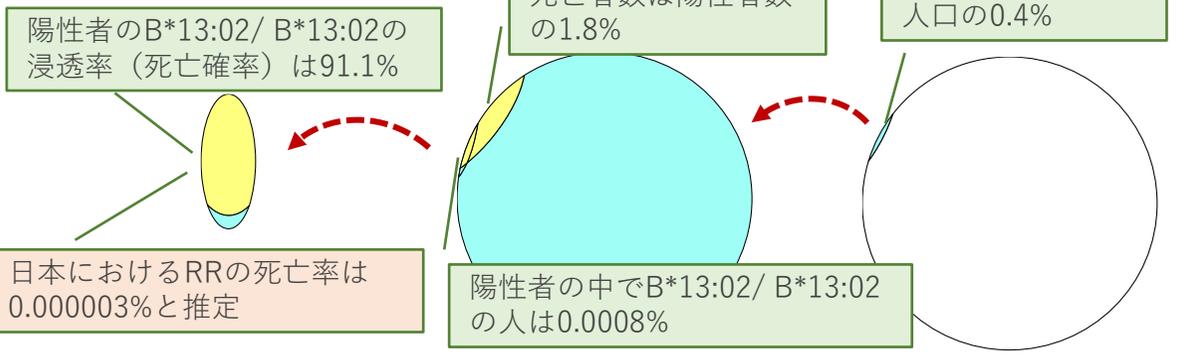
1つの国について複数のDBが存在する場合は、サンプルサイズを重みとした加重平均を取り、国別のアレル頻度とした。

スイスの場合、8つのデータベースがある

データベース	アレル頻度	サンプル数
Switzerland Aargau-Solothurn	0.0202	1,838
Switzerland Basel	0.0291	1,888
Switzerland Bern	0.0302	3,545
Switzerland Geneva pop 2	0.0053	1,267
Switzerland Lugano	0.0269	1,169
Switzerland Luzern	0.0258	1,553
Switzerland St Gallen	0.0276	2,113
Switzerland Zurich	0.0227	4,875

## Point 2

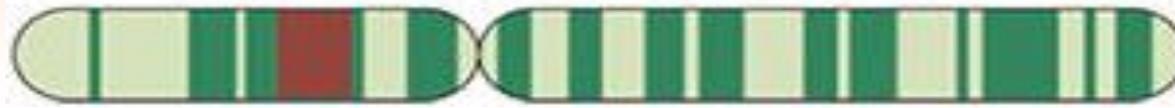
遺伝型RRの浸透率は91.1%であるが、RRは一般集団において頻度が非常に小さいため、実際の死亡率は小さい。たとえば日本であればRRにおける死亡率は $3.12 \times 10^{-8}$ 程度である



# 先行研究との比較

我々の研究で新型コロナウイルス感染による死亡リスクとの関連性が示唆されたHLAアレルにおいて、いくつか先行研究が存在した

6番染色体



HLA-A

HLA-C

HLA-B

- **A\*02:07** … 日本人の53人の重症患者53人と137人の非重症患者を対象とした研究  
重症患者群におけるアレル頻度は非重症患者に比べて有意に高かった  
OR : 2.26; 95%CI : 1.27-3.91; P= 0.0002  
Ishii, T. Cureus.(2020).
- **C\*01:02** … インシリコでの研究  
SARS-CoV-2に対して予測されるペプチド結合が少ないことが明らかになり、重症化リスクが高まる可能性があることを示唆  
Novelli, A. et al. HLA 94, 610-614(2020)
- **C\*08:01** … 中国人患者82人を対象とした研究  
C\*08:01はリスクアレルである可能性を示唆  
(OR : 1.65; 95%CI : 1.05-3.58; P=0.039)  
Wang, W. HLA96, 194-196 (2020).
- **B\*13:02** … 中国人の262人の重症患者と70人の非重症患者を対象とした研究  
アレル頻度は重症患者群で0.058, 非重症患者群で0.051で傾向が示唆された(P=0.049).  
Wang, F. et al. Cell Discov.6, 83 (2020).
- **B\*46:01** … インシリコでの研究  
SARS-CoV-2に対して予測されるペプチド結合が最も少ないことが明らかになり、重症化リスクが高まる可能性があることを示唆  
Nguyen, A. et al. Journal of Virology94 (2020).

## まとめ

- SARS-CoV-2感染の死亡リスクとなるHLAアレルを同定  
**HLA-B\*13:02, A\*02:07, A\*11:01, DQB1\*05:02, B\*46:01, C\*01:02, C\*08:01, DRB1\*09:01**
- その中で  
**HLA-B\*13:02, A\*11:01, B\*46:01, C\*01:02, C\*08:01**  
はSARS-CoV-2 感染による重症化リスクに関連していることを示唆する先行研究が存在
- 本研究の限界
  - ✓ オープンデータベースを用いて解析しているため、データに誤りがあった場合正しい結果が得られない可能性がある
  - ✓ HLAのアレル頻度データベースが存在しない国は解析できない
- SARS-CoV-2感染による遺伝的重症化リスクを推定する数理モデルを構築
  - ✓ 患者データを使用せず全世界のデータを統合的に扱うことができるモデル
  - ✓ 世界中に蔓延する感染症が発生した際にオープンデータから遺伝的リスクを推定できるモデル

