

CAGI slide pack guidelines

I agree to my slides being used according to the following terms: **Rafael Guerrero**

- Authors: please include THIS slide at the beginning of your presentation to indicate that you agree with reuse of your slides in CAGI presentations, according to the terms below, and digitally sign by writing your name above.
- To disseminate the CAGI results, I encourage all CAGI participants to remix these CAGI slides for inclusion in their oral and poster presentations
- CAGI participants may use or publish any content of these slides only in compliance with the CAGI Data Use Agreement
- CAGI participants may not cite or publish any content of these slides except with the written permission of the originating author or in the primary CAGI publications
- CAGI participants must acknowledge the original authors of these slides and CAGI 2016. Include the acknowledgement banner across the bottom
- CAGI participants must include the CAGI credits slide in their presentations
- Slides with a red slash (no remix) may not be included in a presentation unless all attendees are registered CAGI participants who signed the data use agreement
- Slides with a red X over the entire slide may not be used in any circumstances
- Slides with a blue slash (embargoed) may not be used unless the embargo is explicitly lifted
- Slides without slash or X may be tweeted unless flagged with the “No tweet” icon.
- CAGI participants using these slides must explain the “No tweet” icon, which means that the slide content should not be disseminated outside the presenting conference hall. If the presentation venue permits pervasive tweeting, you may not include these slides in your talk 
- Note for slide authors: we expect all CAGI participants to abide by these restrictions, and will make best effort to ensure they are followed. However, but bear in mind that we have limited means of enforcing them, and therefore the restrictions cannot be guaranteed.

SickKids5

Mooney-Radivojac Group

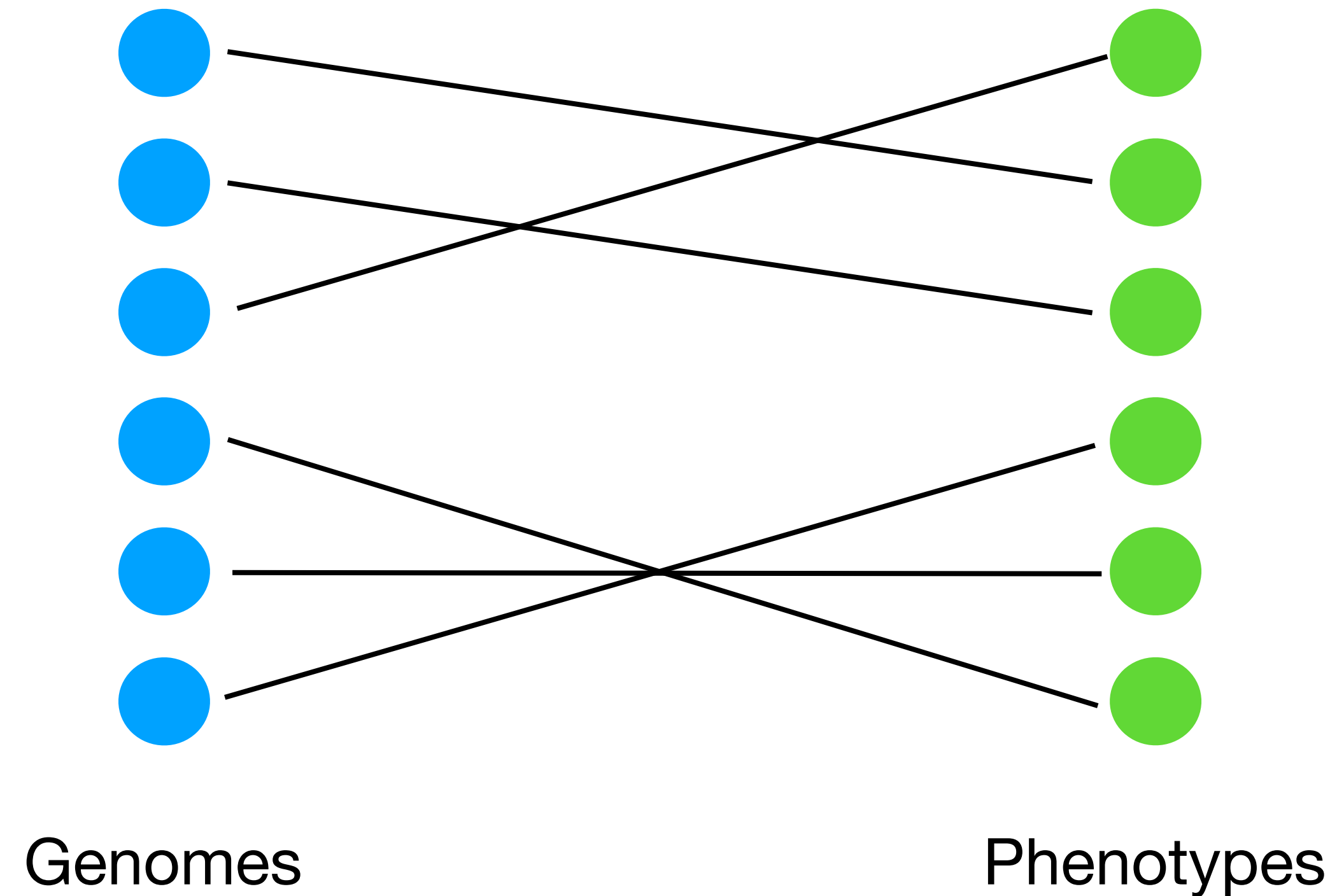


Genomes

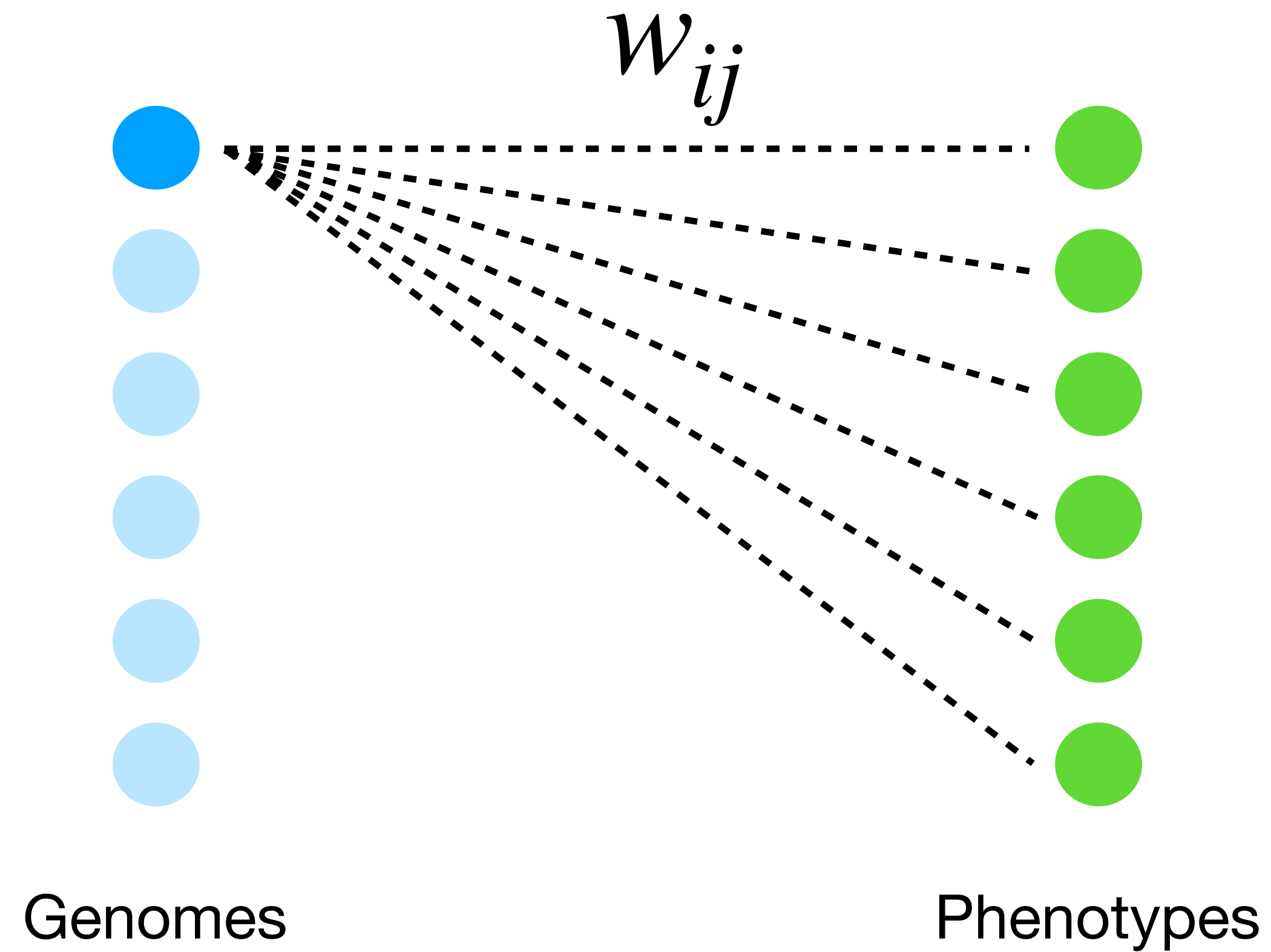


Phenotypes

Bipartite matching

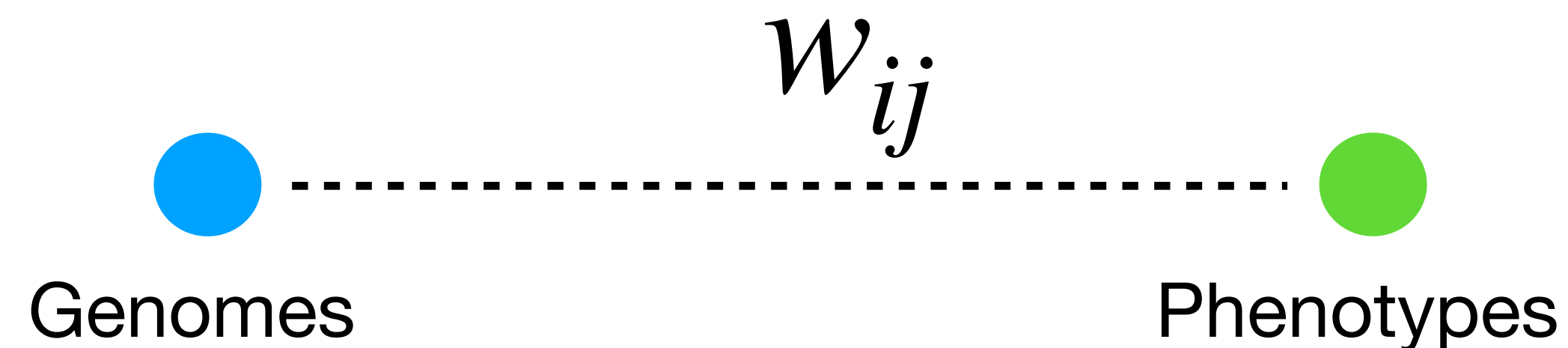


Edge scores



Edge scores

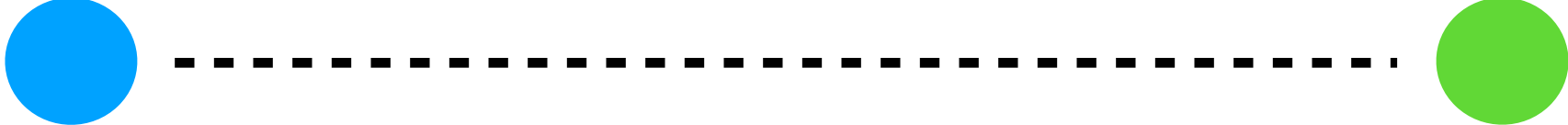
$$w_{ij} = \overset{\text{Sex}}{s(i,j)} \overset{\text{Ancestry}}{(a(i,j))} + \overset{\text{Disease}}{d(i,j))}$$



Sex score component

Does the inferred sex of genome i
match observed sex of j ?

Sex of genome inferred by presence of Y-chromosome variants

$$w_{ij} = \overset{0/1}{s(i,j)}(a(i,j) + d(i,j))$$


Genomes Phenotypes

Ancestry score component

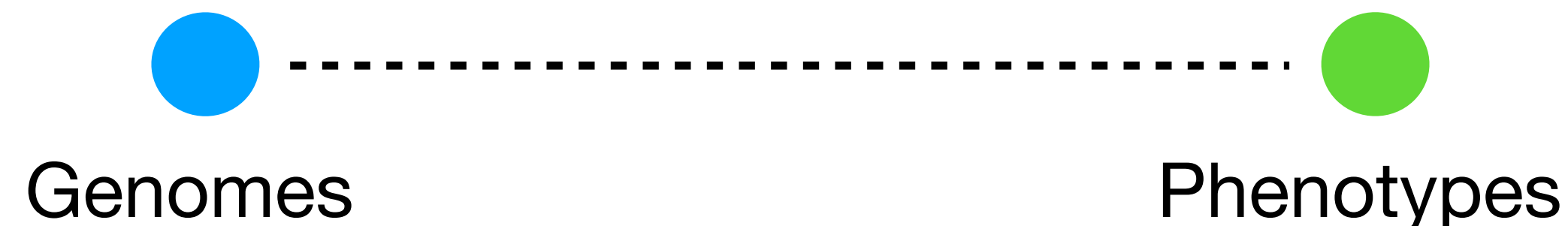
Similarity of genome i and reported ancestries of j

$$a(i, j) = \frac{\overset{\text{EAS}}{\text{mom}} + \overset{\text{EUR}}{\text{dad}}}{2}$$

Ancestries approximated to populations in the 1000G

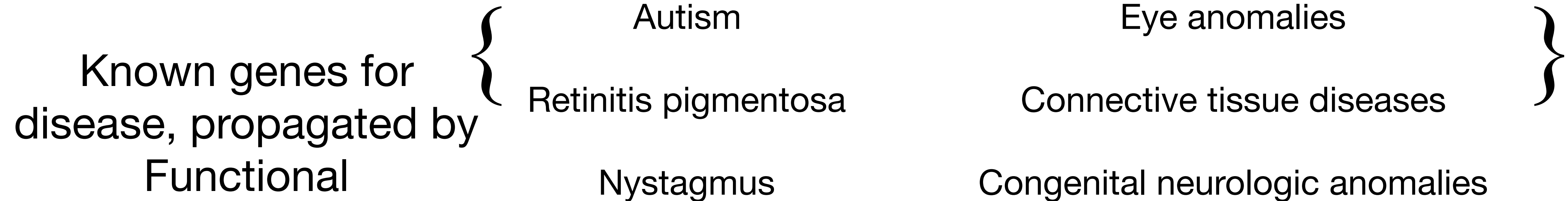
Similarity is 'identity-by-state', relative to other genomes in set

$$w_{ij} = s(i, j)(a(i, j) + d(i, j))$$



Disease score component

Gene sets



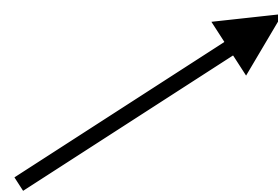
(Nabieva & Singh, 2005)

$$w_{ij} = s(i,j)(a(i,j) + d(i,j))$$

Genomes

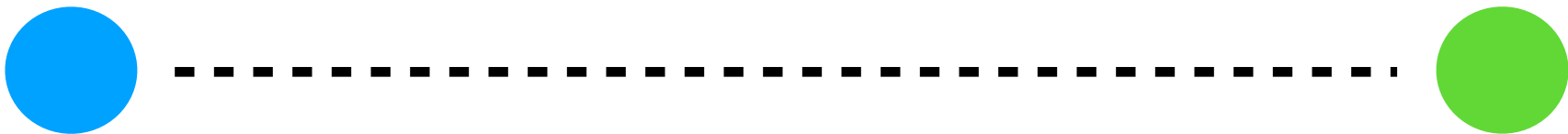
Phenotypes

Disease score component

Gene sets  **Uncommon variants in set (MAF<0.1)**
Evaluated with:

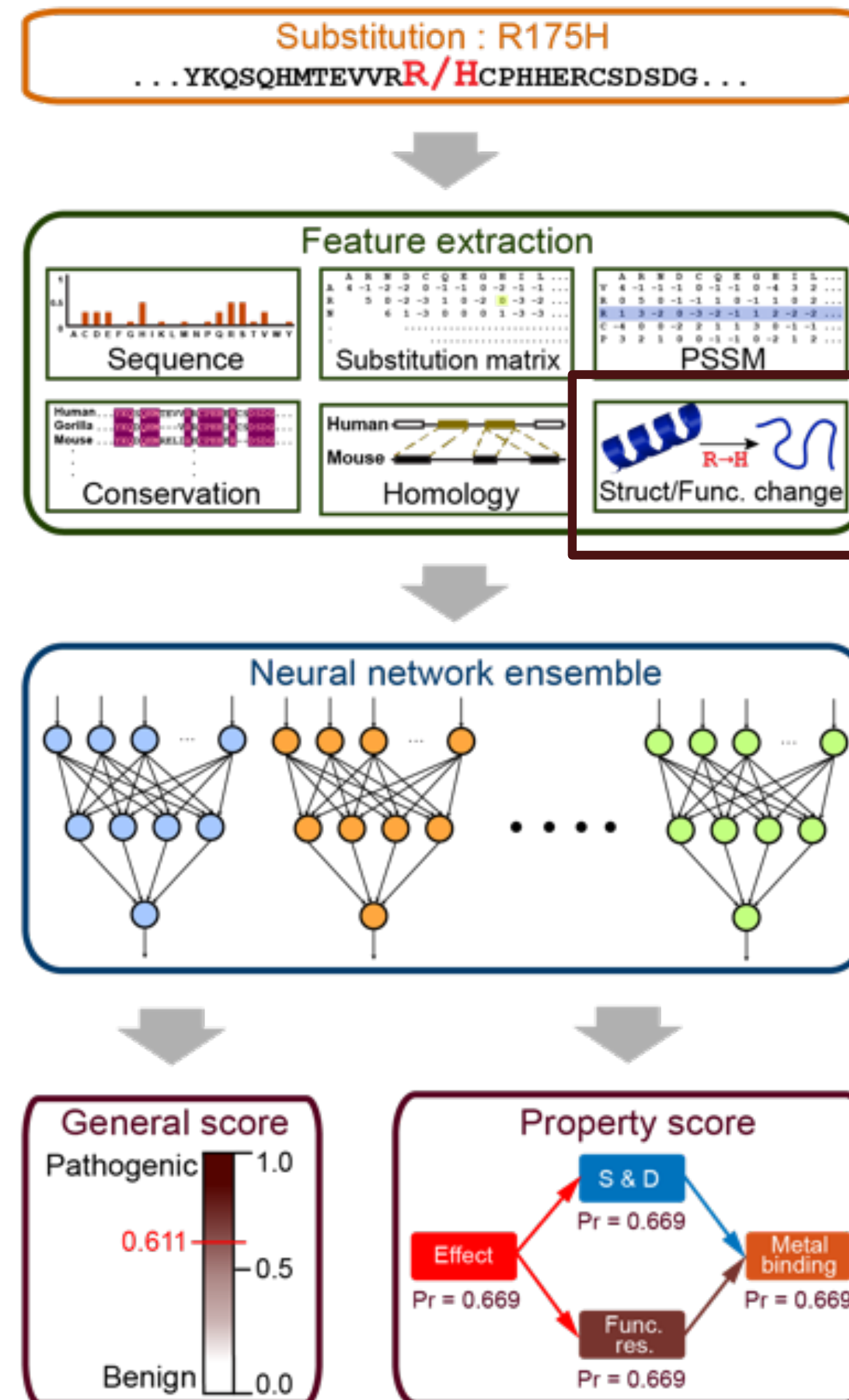
MutPred2 (Pejaver et al 2017) — Missense

MutPred-LOF (Pagel et al 2017) — Frame-shifts, Stop-gains

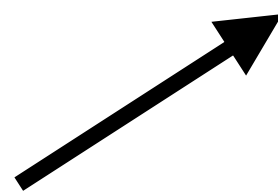
$$w_{ij} = s(i,j)(a(i,j) + d(i,j))$$


Genomes Phenotypes

MutPred2 and MutPred-LOF



Disease score component

Gene sets  **Variants in set**
Get variant with max score per gene

P-value for gene from $B(\alpha, \beta)$

Total score is sum of log(P-values)

Parameters from
distribution of
MutPred scores of
gnomAD variants

$$w_{ij} = s(i, j)(a(i, j) + d(i, j))$$



Genomes



Phenotypes

Disease score component

$$d(i,j) = q \left(\begin{array}{c} \text{Score for} \\ \text{Primary} \\ \text{Disease} \end{array} \right) + (1-q) \left(\begin{array}{c} \text{Score for} \\ \text{Secondary} \\ \text{Disease} \end{array} \right)$$

Retinitis

Nystagmus

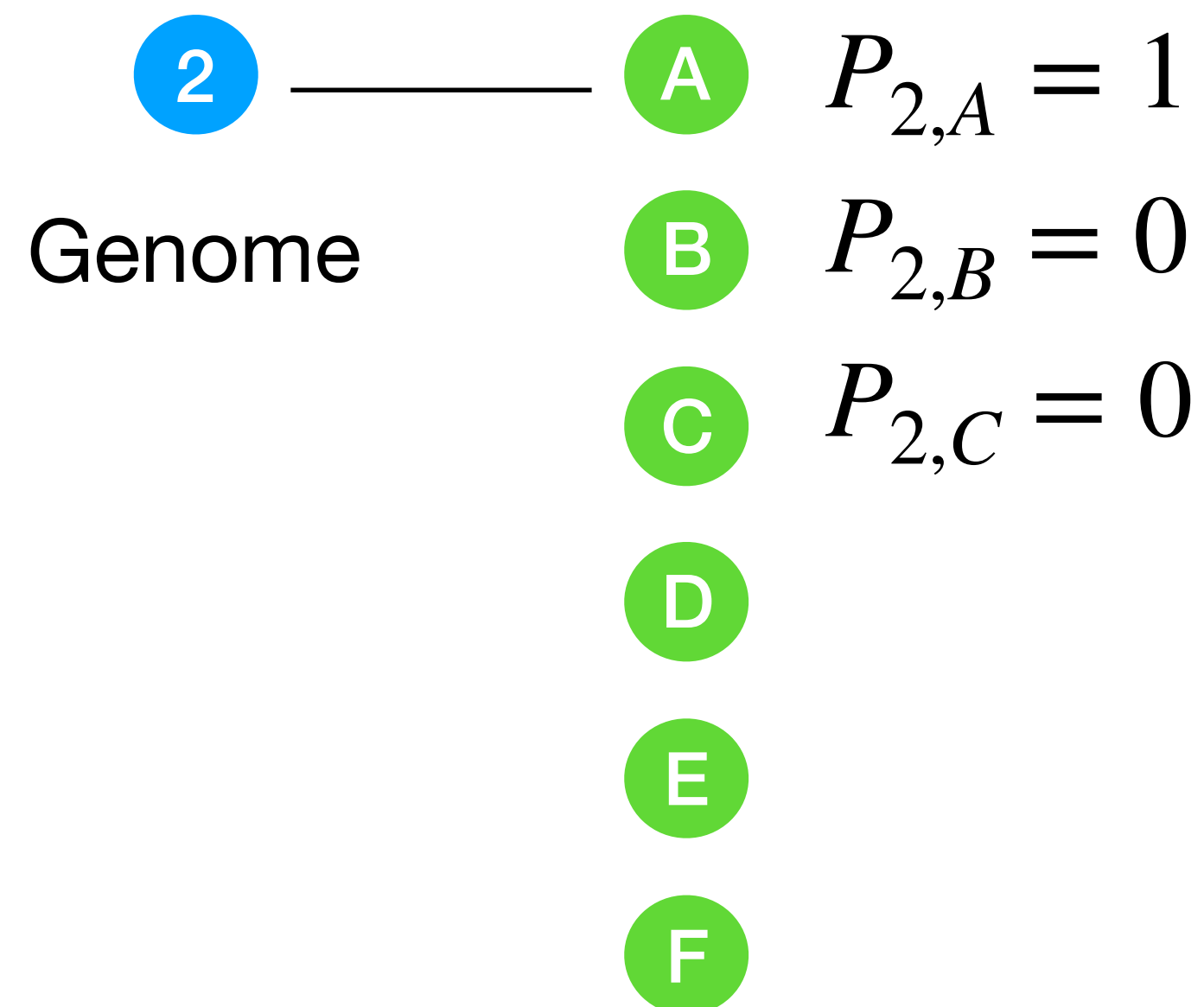
$$w_{ij} = s(i,j)(a(i,j) + d(i,j))$$

Genomes Phenotypes

$$q \sim U(0,0.3)$$

Probability of assignment

Result of one
bipartite matching



Repeat, varying weights

$$w_{ij} = s(i, j) + (a(i, j)p + d(i, j)(1 - p))$$

$$p \sim U(0.2, 0.8)$$

Compute proportion

$$P_{2,A} = \frac{n_{2,A}}{N} = \frac{565}{1000}$$

Acknowledgments

Pedja Radivojac

Sean Mooney

Vikas Pejaver

Kymberleigh Pagel

Yuxiang Jiang