PIRCHE

Anonymization of HLA genotypes for communication with untrusted parties

Matthias Niemann, VP Technology, PIRCHE AG

Agenda

What is HLA why is it important in organ transplantation

Why to be careful when sharing HLA data

Specific attacks on HLA (and genomic) data

The bigger picture

NEPHRO CAGE

Supported by:

Federal Ministry for Economic Affairs and Climate Action

on the basis of a decision by the German Bundestag

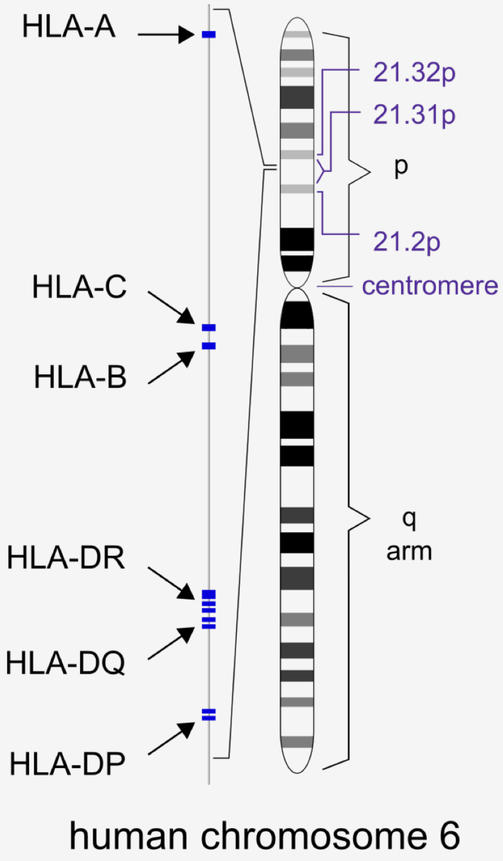
Humans are 99.9% genetically identical.

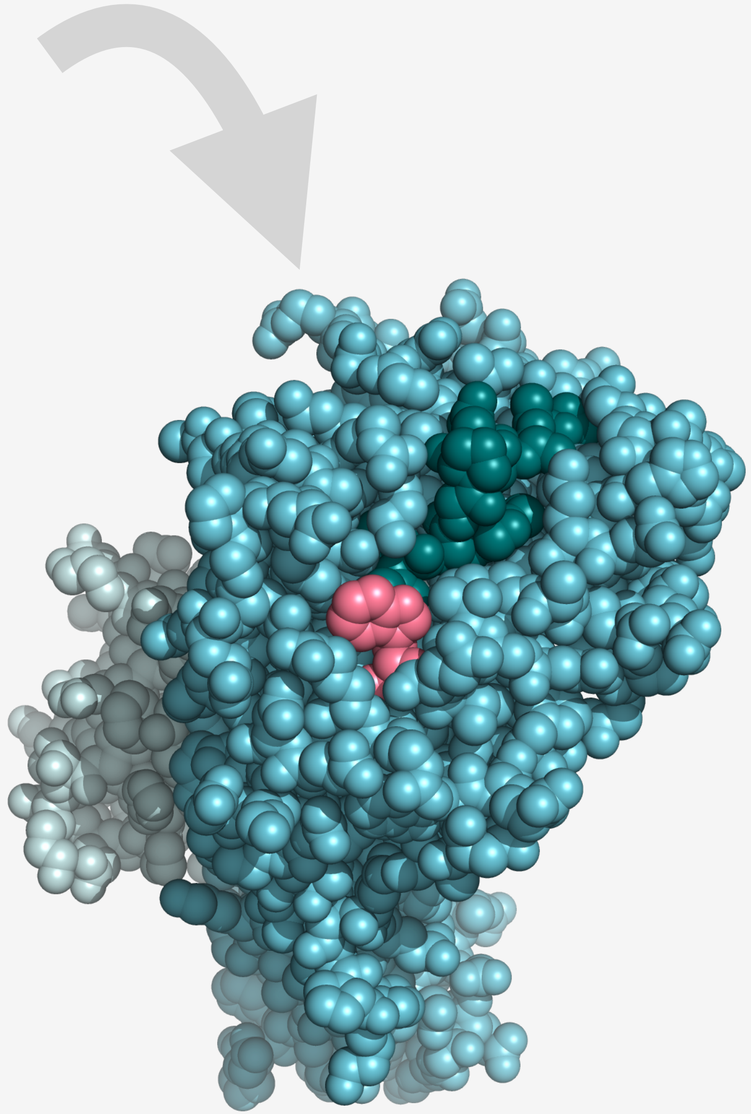
So what is in the 0.01%?

PIRCHE

What is HLA?

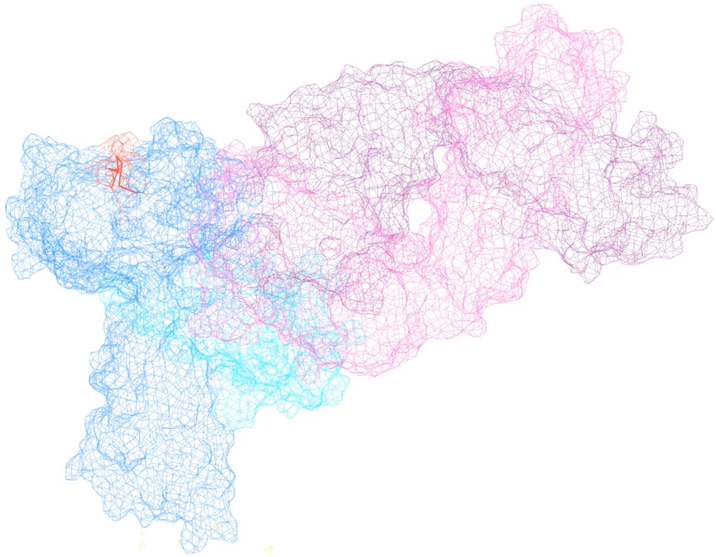
- Human genome encodes Leukocyte Antigen (HLA) loci
- HLA are transcribed into cell surface proteins responsible for immune response
- Evolutionary benefit for species to have highly variable HLA gene region
- 2 x 6 genes per individual (plus a couple more)

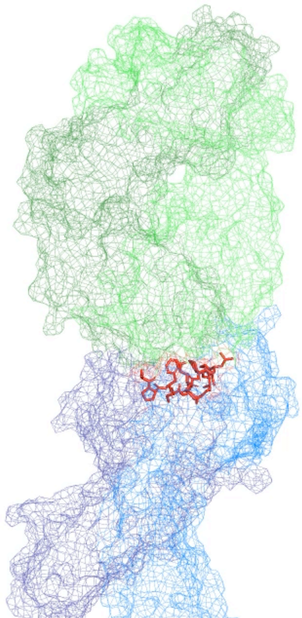




PIRCHE

Why is HLA important?





- HLA defines our ability to adapt to diseases
- HLA is inherited
- HLA incompatibility is a major problem for transplantation
- "Compatibility" is a complex problem

PIRCHE

What we (don't) know about HLA compatibility

- Identical proteins (i.e. sequences) are accepted
- (Some) amino acid differences on the protein surface are targets for antibodies, others trigger cytotoxic T cell responses
- (Some) amino acid differences in the whole protein cause T helper responses
- T helper cells support antibody formation

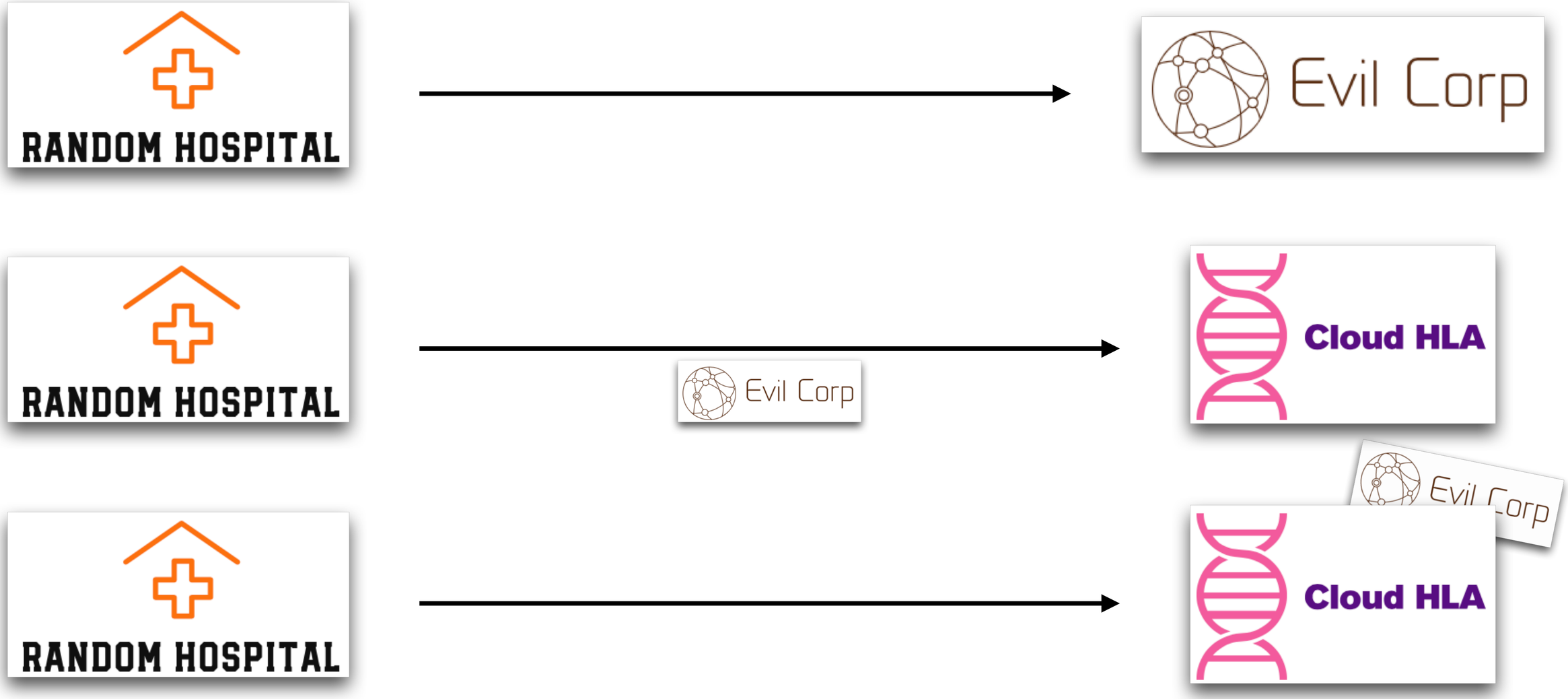
How that's tackled with computers

- Bioinformatic prediction pipelines to
  - predict protein folding
  - characterise protein surface
  - predict protein interaction (peptide binding/docking)
- Biostatistics
- Cloud computing
  - (...and that's the PIRCHE product)

Legal implications of sharing HLA data

- HLA data of a patient is not obvious, yet a powerful composite identifier
- HIPAA considers HLA data as de-identified (i.e. not protected health information)
- PIPEDA requires “no serious possibility” of re-identification
- GDPR considers HLA as pseudonymized
- Anonymized data not in the scope of GDPR
- But it's not only about legal...

What if...



Is it a thing?

- **Neighbor told me...**
  - **he‘s on dialysis, waiting for kidney transplantation**
  - **mother is Japanese, father African American**

HLA haplotype frequency tables

- **Listening to incoming HLA data may allow to...**
  - **learn when an organ offer was made to my neighbor**
  - **learn about the HLA typing of the donor**
  - **learn about ethnicity of the donor**
  - **learn about history of the donor**

IP address/user, time of request

Look up donor in haplotype frequencies

Patient donor linkage

Local news articles

PIRCHE

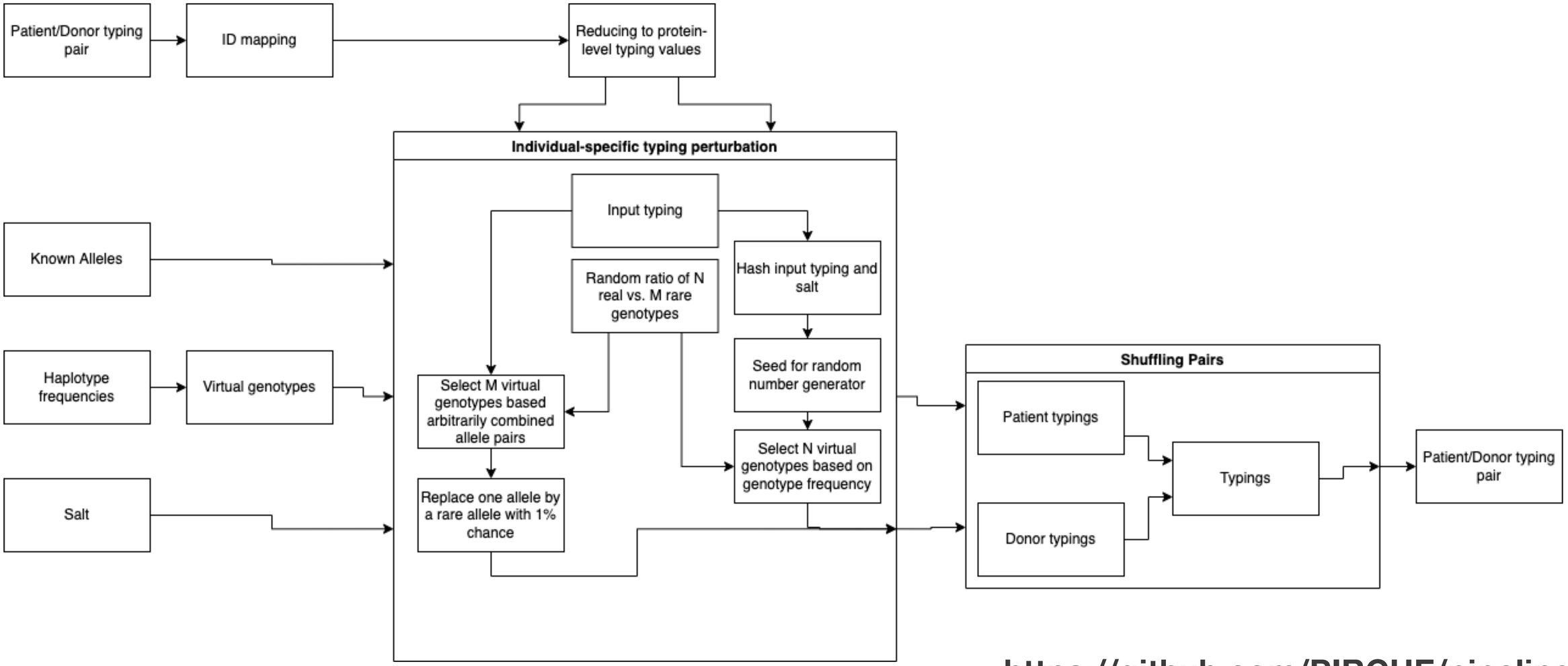
Anonymization strategy

- Binning data "destroys" information
  - Not desired when applying highly sensitive prediction method
- Perturbation "buries" the real data in counterfeit records

The risk of "just some random data"

- **Attack 1 – domain-specific knowledge attack: HLA follows linkage disequilibrium**
- **Attack 2 – repeated request attack: filtering random data and remain the real data**
- **Attack 3 – dictionary attack: map all potential HLA genotypes to obfuscated data and reverse the inputs**
- **Attack 4 – family donor attack: repeated requests may indicate living donation, extract overlapping haplotypes as patients are probably related**
- **Attack 5 – typing level knowledge: knowing about the typing methods applied in the lab, certain anonymized values are not plausible**
- **Attack 6 – typing level difference: labs apply different methodologies depending on the transplant setting (living/deceased, historic vs. current)**

Anonymization client



https://github.com/PIRCHE/pipeline

Conclusions of the use case

- Sharing HLA data with (trusted) vendors is fine by current legislation (HIPAA, PIPEDA and GDPR)
- Currently no public datasets available to map HLA

- But...
  - there is value to extract from pseudonymized data
  - data can be aggregated with certain assumptions
  - more data sources may become available

Conclusions

- Think thoroughly about the data flow
- "Allowed" is not "accepted"
- Avoiding a leak of data requires diving into domain-specificities

- Pirche is hiring! Feel free to reach out to matthias.niemann@pirche.com

PIRCHE