Anonymization of HLA genotypes for communication with untrusted parties

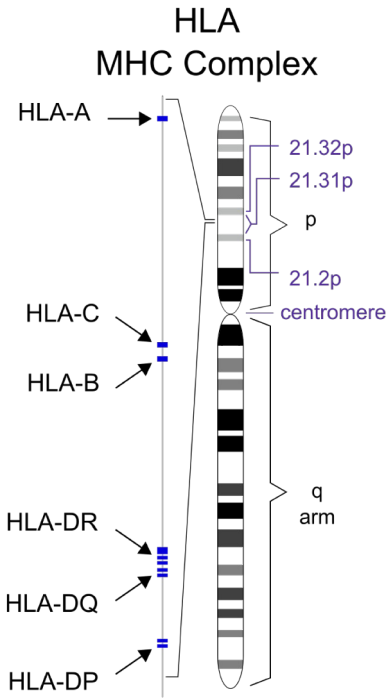
Matthias Niemann
Vice President Technology
PIRCHE AG

Photo by Elf Moondance

NEPHROCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
https://nephrocage.org

The HLA System

- 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



human chromosome 6
wikipedia.org

Anonymization of HLA genotypes for communication with untrusted parties

Matthias Niemann,
VP Technology
PIRCHE AG,
2022-08-16

Chart **2**

Why sharing HLA data?

- In transplantation, mismatched donor-HLA impairs outcome
- Recent technologies refine understanding of "match" or "mismatch" (i.e. epitope matching)
- These technologies are computationally expensive
  - not feasible to run locally at hospital
  - provided as a Cloud service

...so we share HLA data with 3rd parties.

Anonymization of HLA genotypes for communication with untrusted parties

Matthias Niemann,
VP Technology
PIRCHE AG,
2022-08-16

Legal implications

- HLA data of a patient is not obvious
- HLA data is however a powerful composite identifier

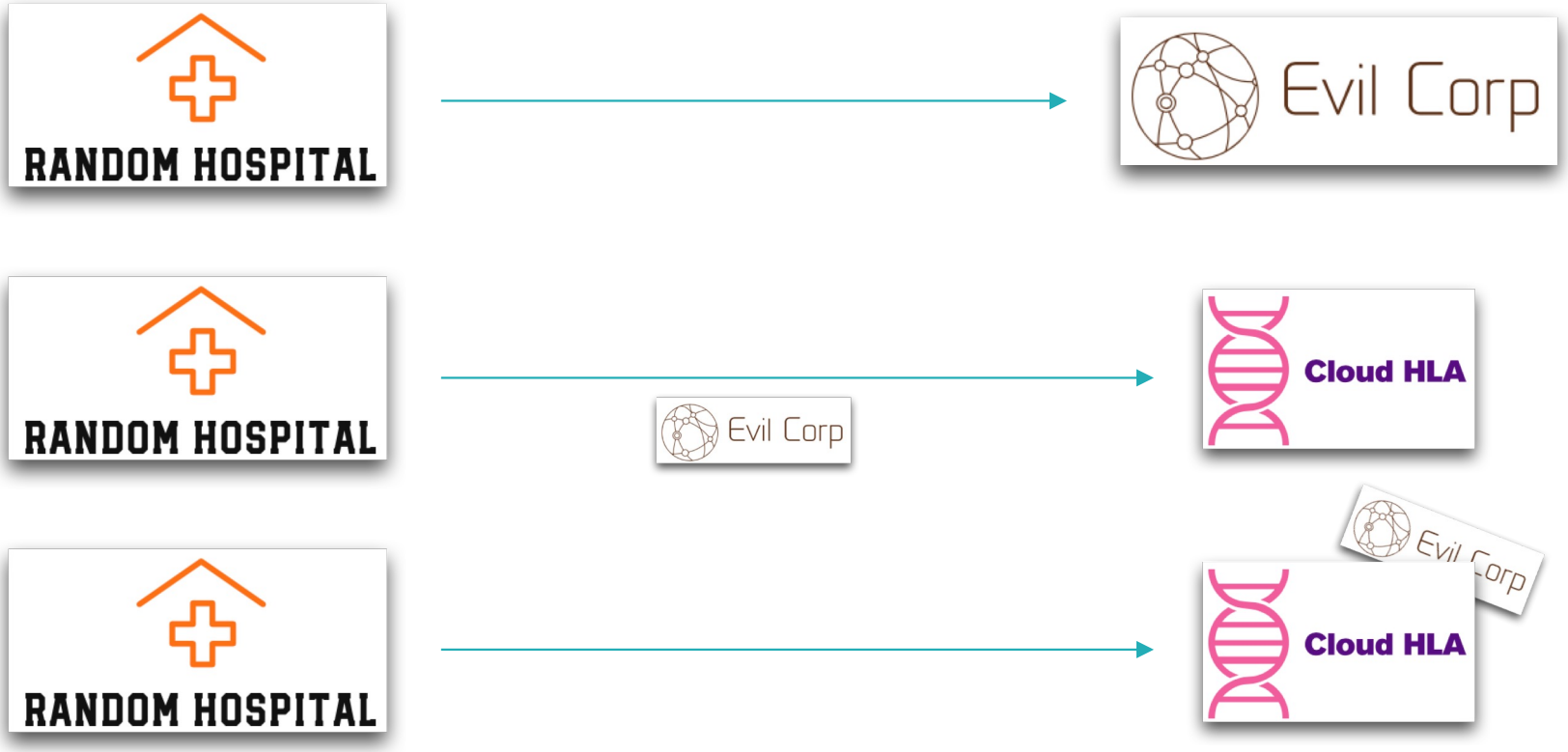
- HIPAA considers HLA data as de-identified (i.e. not protected health information)
- PIPEDA requires “no serious possibility” of re-identification
- GDPR considers HLA as pseudonymized

- Anonymized data not in the scope of GDPR

Anonymization of HLA genotypes for communication with untrusted parties

Matthias Niemann,
VP Technology
PIRCHE AG,
2022-08-16

But what if...



Anonymization of HLA genotypes for communication with untrusted parties

Matthias Niemann,
VP Technology
PIRCHE AG,
2022-08-16

Is that a thing? Example

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

HLA haplotye 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

Aggregate with local news articles

Anonymization of HLA genotypes for communication with untrusted parties

Matthias Niemann,
VP Technology
PIRCHE AG,
2022-08-16

NEPHROCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
https://nephrocage.org

Anonymization Strategy

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

Anonymization of HLA genotypes for communication with untrusted parties

Matthias Niemann,
VP Technology
PIRCHE AG,
2022-08-16

„Let‘s produce 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)

Conclusion

- 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

Matthias Niemann,
VP Technology
PIRCHE AG,
2022-08-16