GenMatch: Secure DNA Compatibility Testing

M. Sadegh Riazi*, Neeraj K. R. Dantu*, L. N. Vinay Gattu*, Farinaz Koushanfar†

***Rice University, Houston, TX, USA**

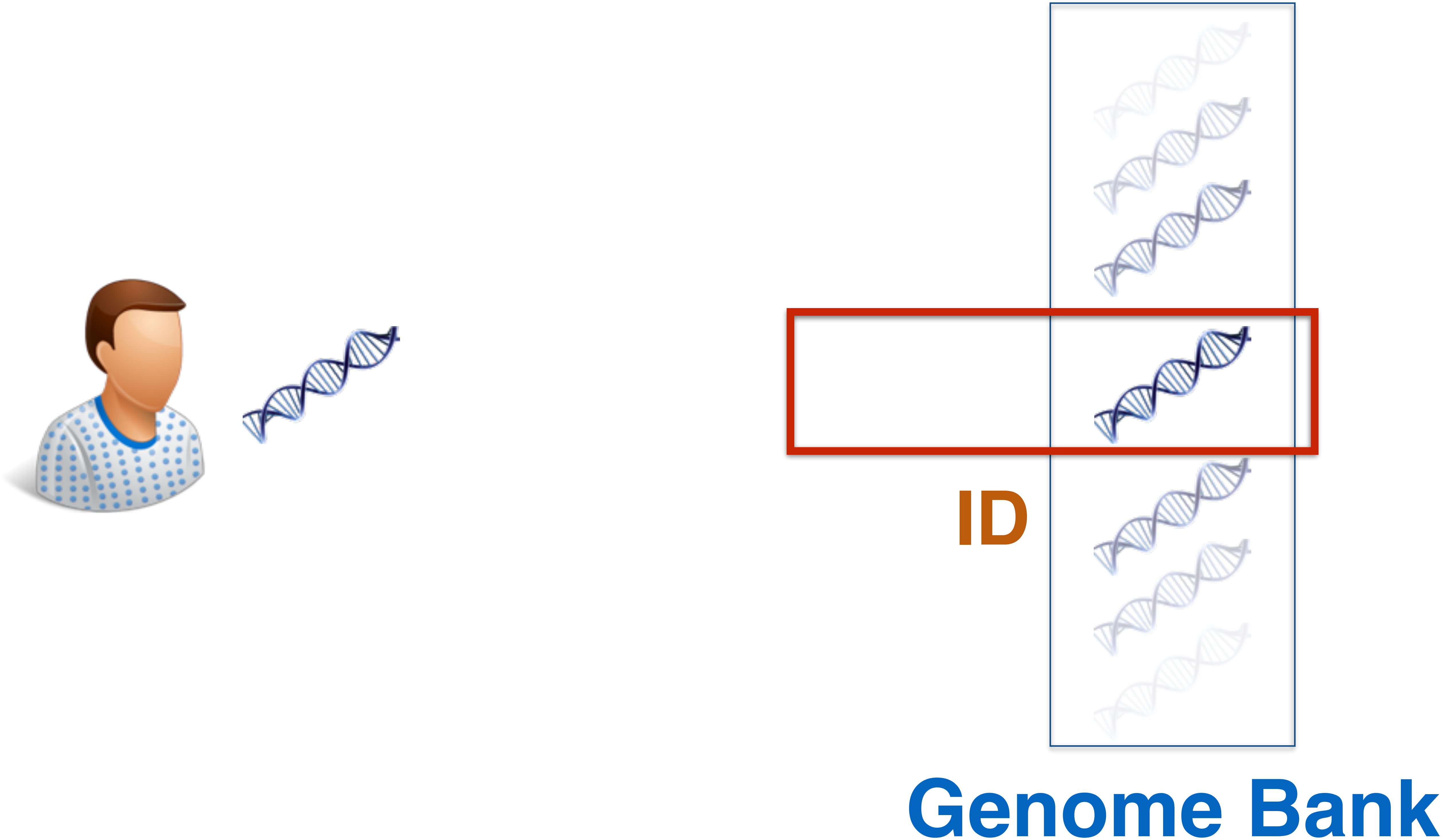
†University of California, San Diego, USA

IEEE International Symposium on Hardware Oriented Security and Trust (HOST), May 5th, 2016

Scenario

- Human Leukocyte Antigen (HLA) analysis which is a crucial test in organ transplantation
- Patient holds her whole genome sequence
- Genome information of all donors are stored in hospital's bank

Current Process

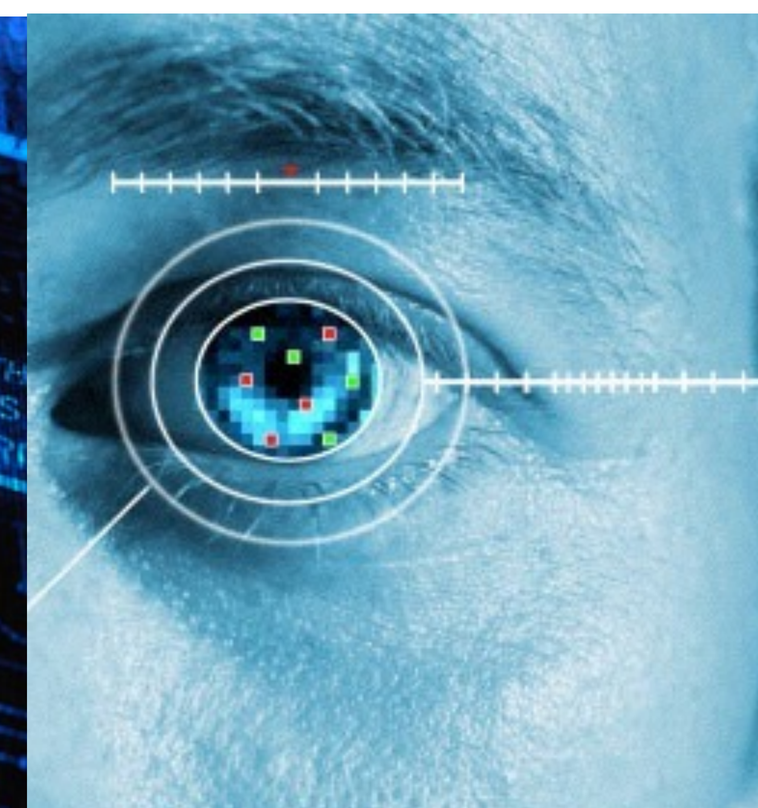


RICE

Why Privacy is Crucial?

- Genome represents ultimate biological identity
- Reflects information about predisposition to a specific disease
- Reveals extensive information about relatives and ancestors
- Irreversible, unlike passwords

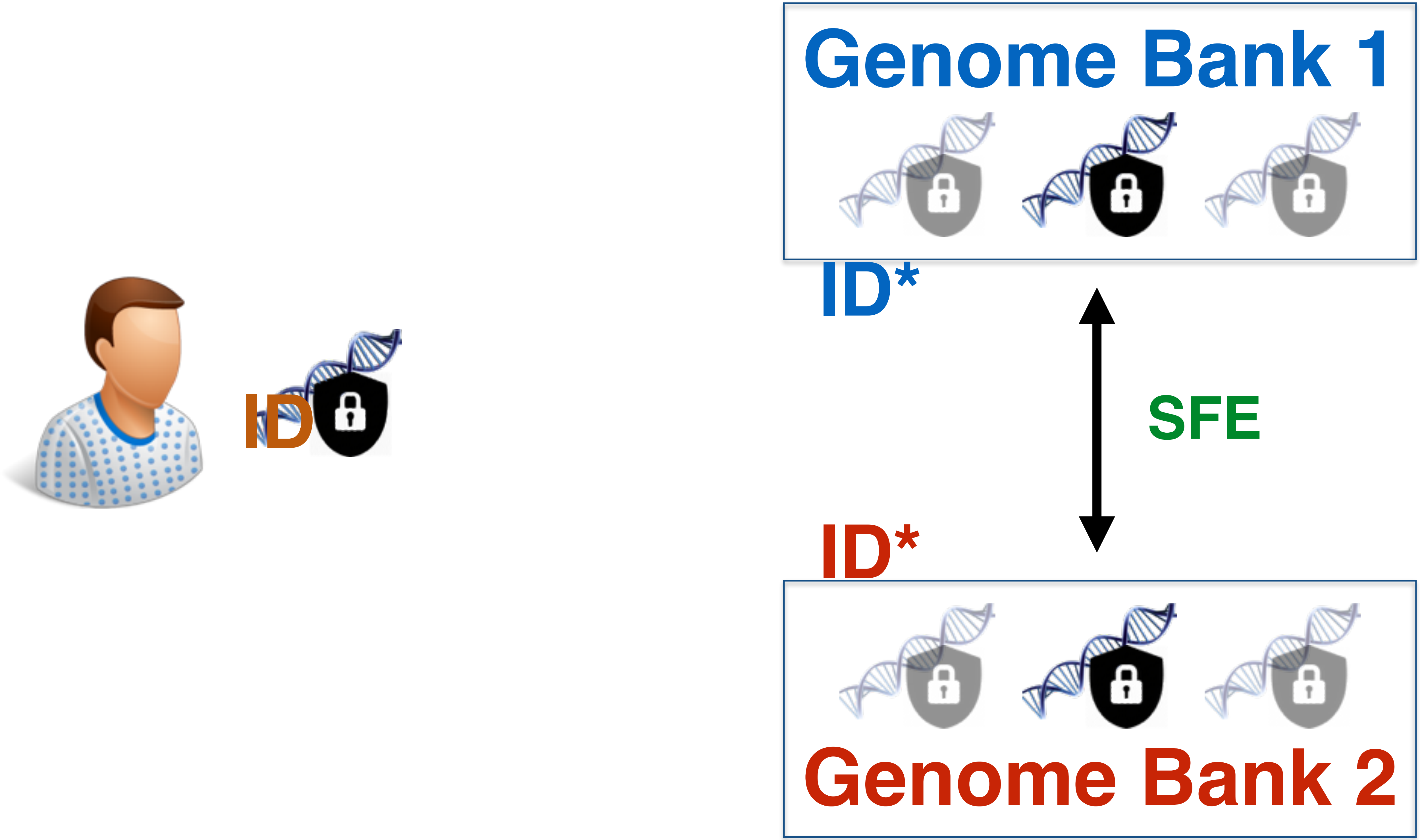




Related Work

- Few works with the same scenario and application
- Cannot scale well due to their approach (mainly based on public key encryption)
- Do not have pre-processing stage and that makes them impractical because they need to securely process gigabytes of DNA data

Our Approach



What is Secure Function Evaluation (SFE)?

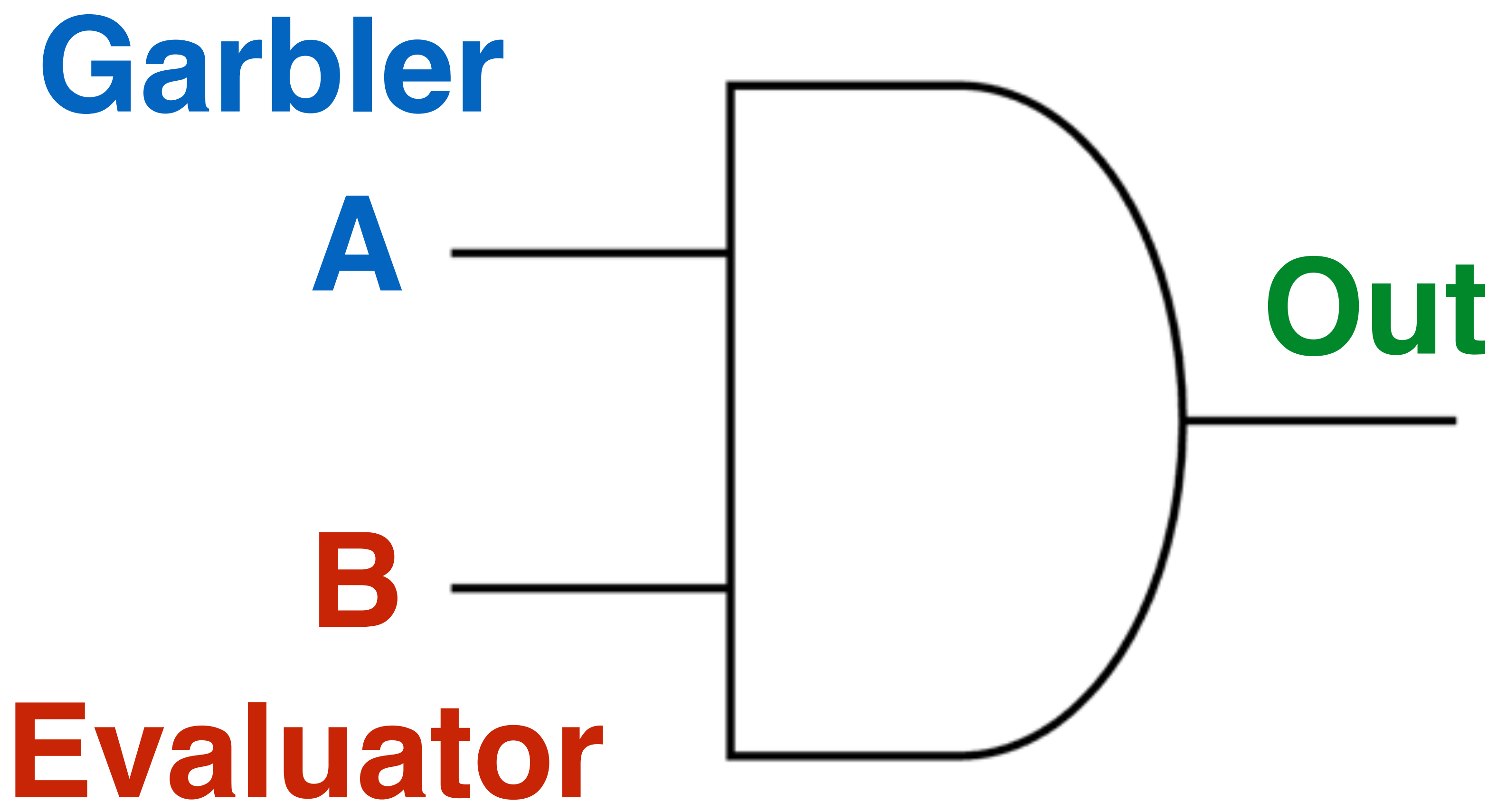
- Two or more parties want to jointly compute a function on their inputs while keeping their own data private
- Example: Yao's millionaire problem
- Is that even possible ?!!

Andrew Chi-Chi Yao 1986:
Any efficiently computable function can be evaluated securely.

Yao’s Garbled Circuit Protocol

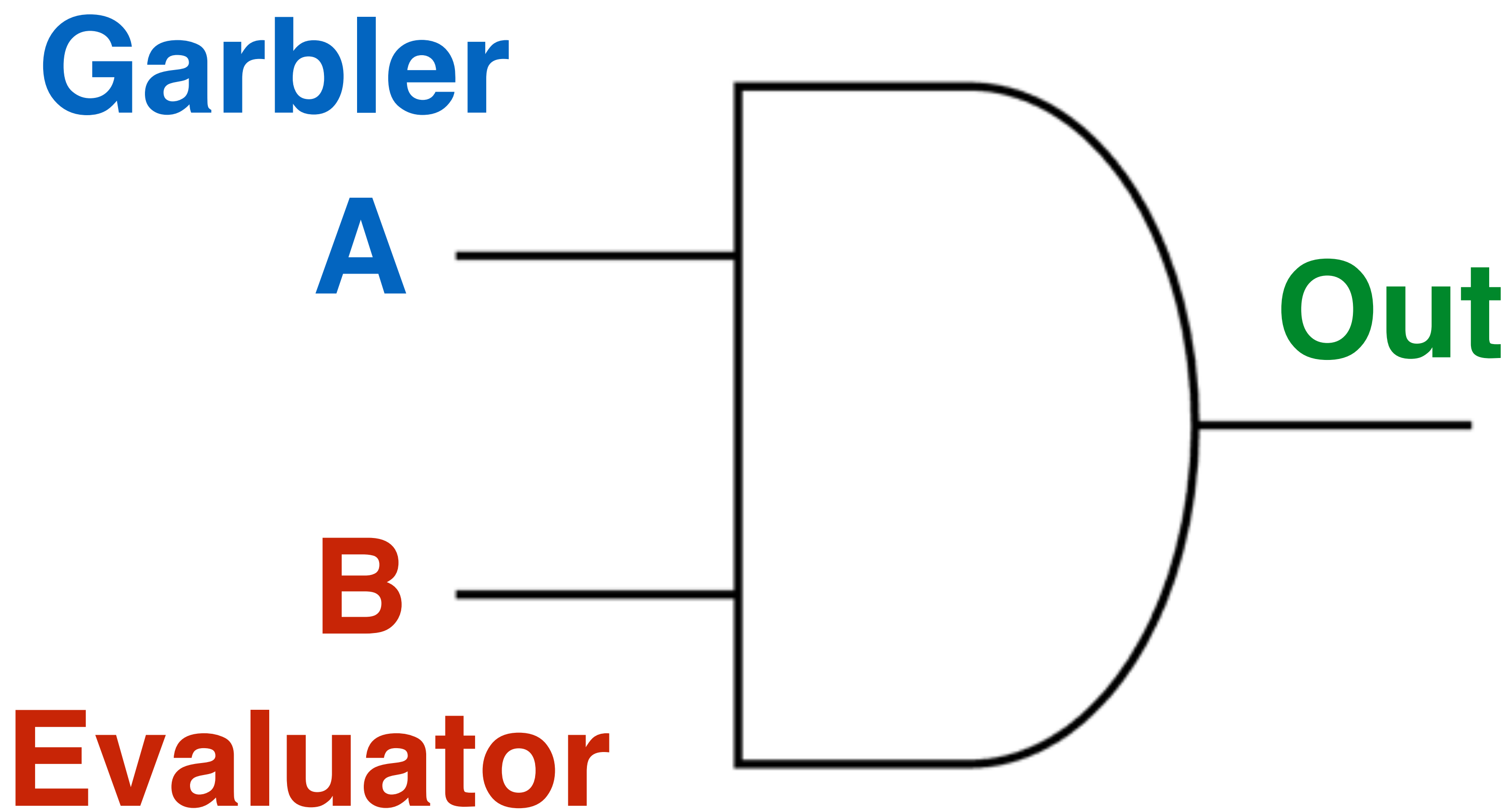
- Describe f(.) in Boolean circuit
- Involves 2 parties (Garbler & Evaluator)
  - Garbling the circuit
  - Sending Information
  - Evaluating the garbled circuit and finding the results
- Secure against honest-but-curious adversary
  What is adversary’s power?

Boolean Gate



A	B	Out
0	0	0
0	1	0
1	0	0
1	1	1

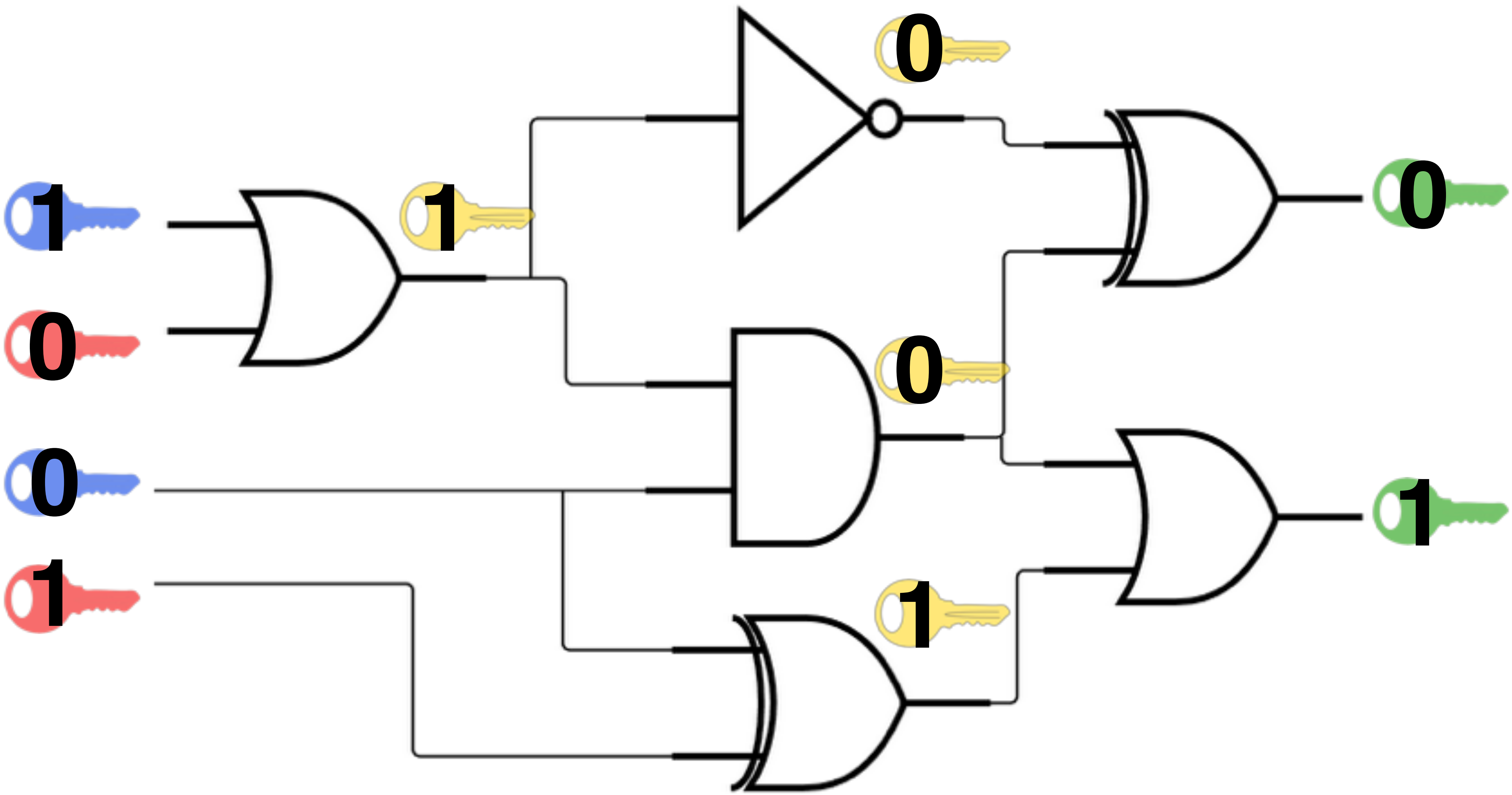
Garbled Gate



A	B	Out
0	0	0
0	1	0
1	0	0
1	1	1

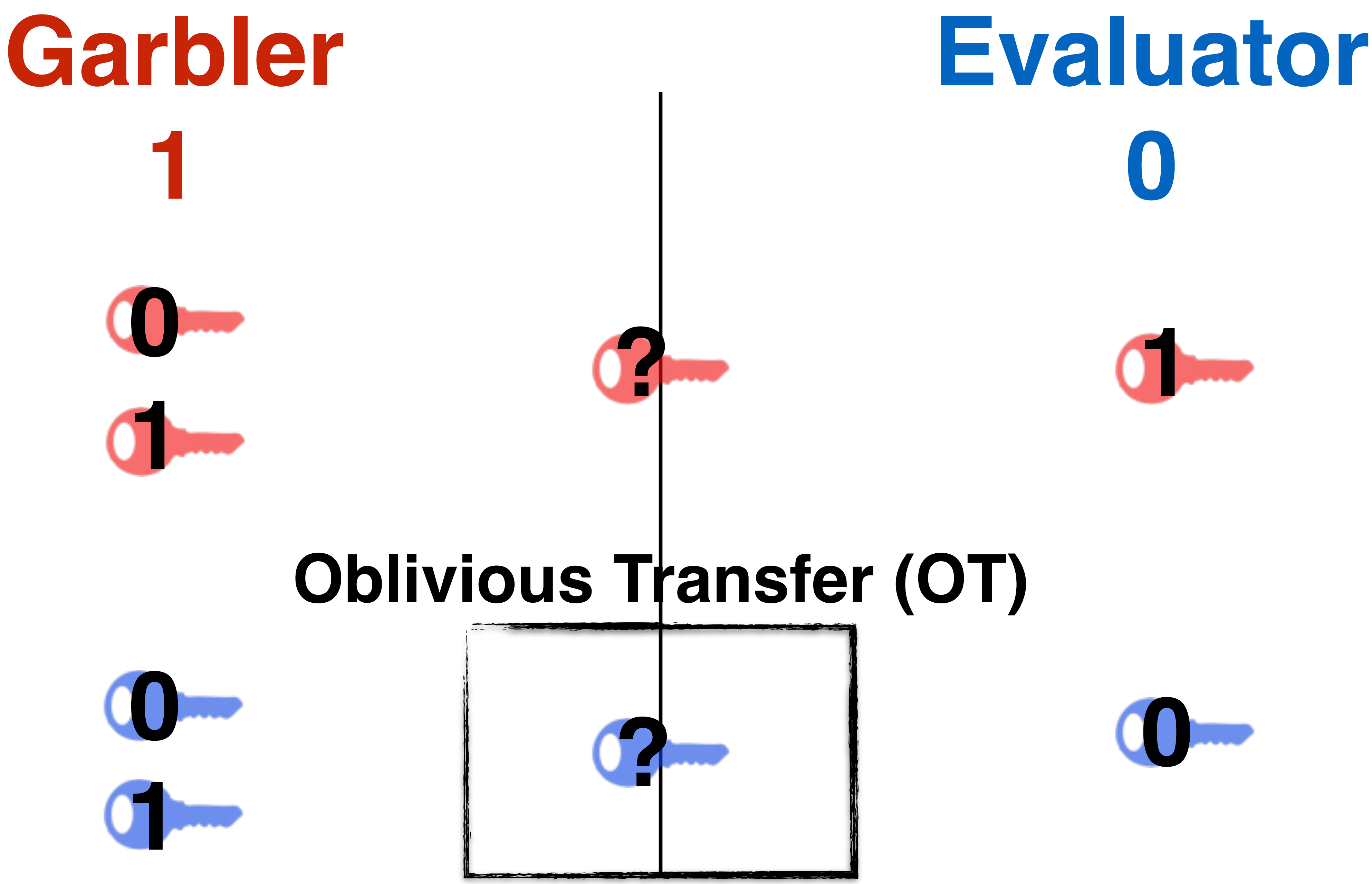
$$(\tilde{\omega}_k^0 \,\|\, \pi_k^0) \oplus H(\tilde{\omega}_i^0 \,\|\, \tilde{\omega}_j^0)$$

Garbled Circuit



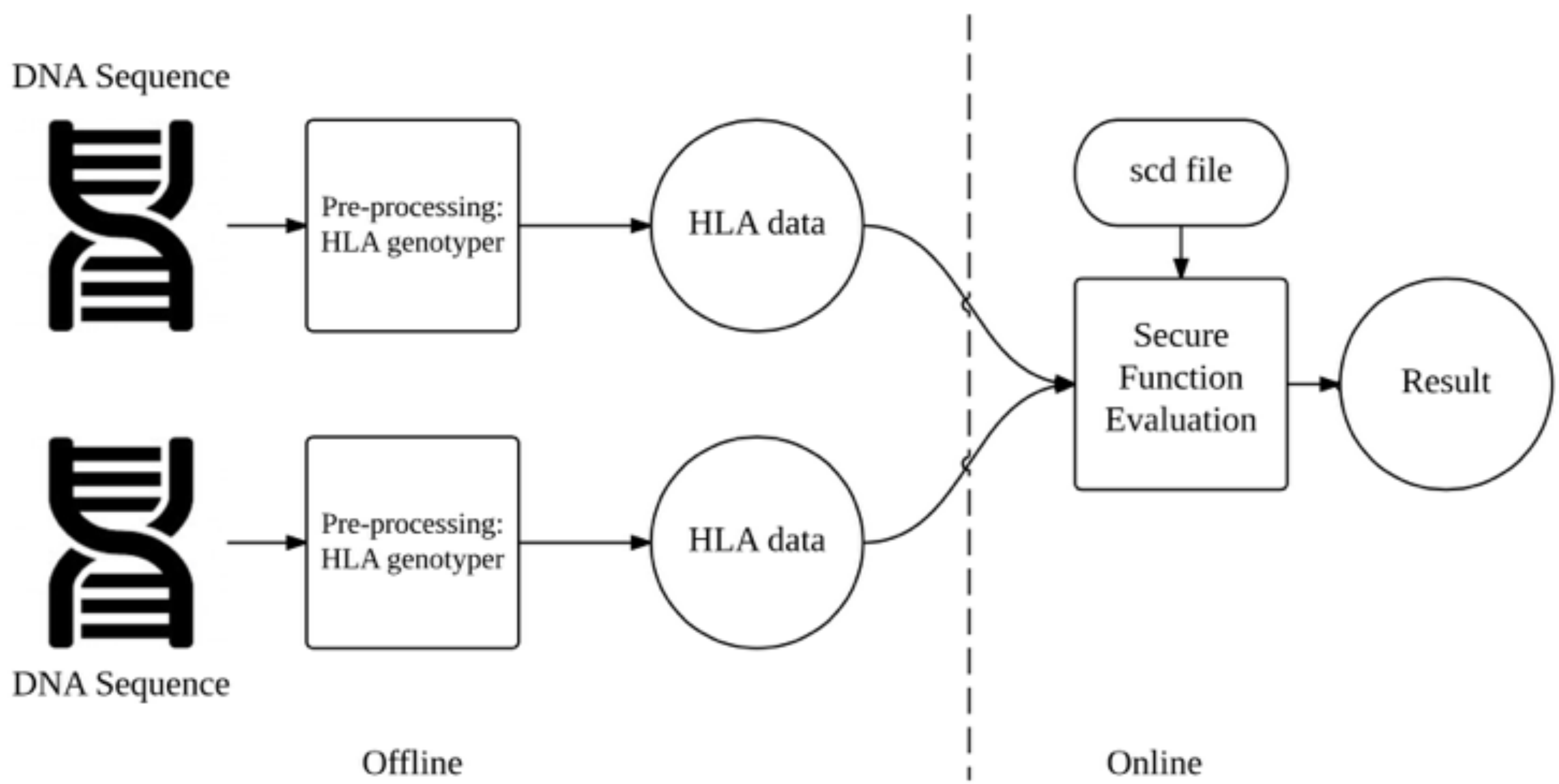
Inputs **Intermediate Values** **Outputs**

Transferring Input Keys



Global Architecture

- **Offline** phase: local and no need for secrecy
- **Online** phase: interactive and secure



RICE

Algorithm

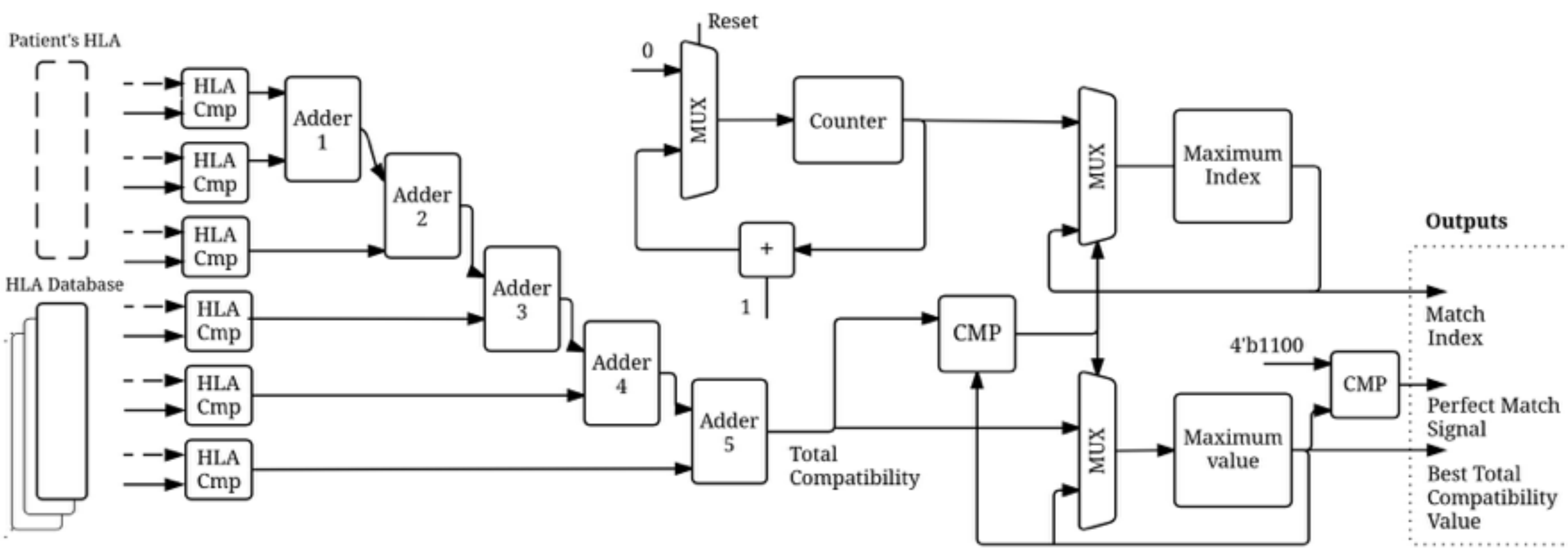
- Each individual holds **6 pairs of HLA data**

- Match and cross match means more compatibility

```
total compatibility = 0
for n = 1 to 6 do
  if HLA1[n][1] == HLA2[n][1] then
    if HLA1[n][2] == HLA2[n][2] then
      compatibility = 1
    else
      compatibility = 0.5
    end if
  else if HLA1[n][2] == HLA2[n][1] then
    if HLA1[n][1] == HLA2[n][2] then
      compatibility = 1
    else
      compatibility = 0.5
    end if
  else if HLA1[n][1] == HLA2[n][2] then
    compatibility = 0.5
  else if HLA1[n][2] == HLA2[n][2] then
    compatibility = 0.5
  else
    compatibility = 0
  end if
  total compatibility = total compatibility + 1/6 × compatibility
end for
```

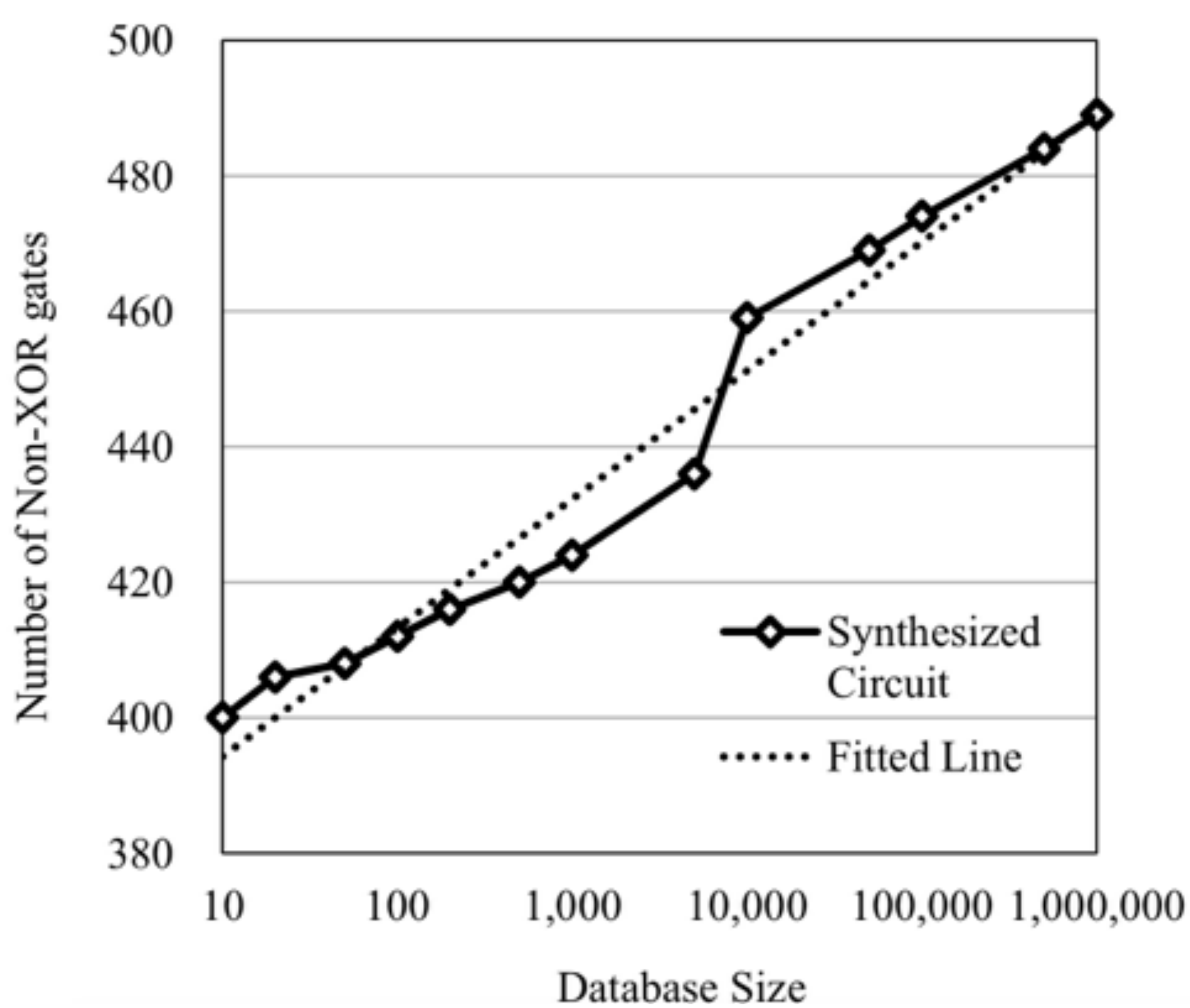
Boolean Circuit

- Several hand-craft optimizations
- Translating floating point operations into integers
- Hierarchical addition structure



Circuit Size Model and Scalability

$$\text{Total \# of garbled tables} = N \times (\alpha \times \log N + \beta)$$



Timing Results

- Securely searching for a compatible DNA in a genome bank of size million, only takes less than two hours!

Database Size	# of XORs	# of Non-XORs	Total Gates	Total Garbled Tables	Communication (MBytes)	Time (s)
10	438	400	838	4,000	1.0	0.07
100	447	412	859	41,200	10.5	0.62
1,000	457	424	881	424,000	108.5	5.79
10,000	433	459	892	4,590,000	1,175.0	63.20
100,000	436	474	910	47,400,000	12,134.4	546.09
1,000,000	439	489	928	489,000,000	125,184.0	5,132.25

Conclusion

- First scalable and efficient GC-based solution for secure organ transplantation compatibility testing
- Designing sub-linear size circuit for HLA compatibility testing
- Proof-of-concept implementation for a bank containing a million encrypted genome profiles

Thank you

Questions?