Membership Inference Against DNA Methylation Databases: Attacks and Defenses

Inken Hagestedt\textsuperscript{1}, Mathias Humbert\textsuperscript{2}, Pascal Berrang\textsuperscript{1}, Irina Lehmann\textsuperscript{3}, Roland Elis\textsuperscript{4}, Michael Backes\textsuperscript{1}, Yang Zhang\textsuperscript{1}

\textsuperscript{1}CISPA Helmholtz Center for Information Security  \textsuperscript{2}Swiss Data Science Center \textsuperscript{3}Helmholtz Centre for Environmental Research Leipzig, UFZ, \textsuperscript{4}German Cancer Research Center, DKFZ

The Question

Are membership inference attacks possible given only mean \( \mu \) and standard deviation \( \sigma \)?

The Data

- DNA methylation: additional molecule (methyl group) attached to DNA
- represented as value in [0, 1]
- methylation patterns vary between tissues, due to environmental factors and due to diseases

Statistics-based Attack

\[ L_\text{test}(x^i) = |x^i - \mu^i|\]

ML-based Attack

\[ f_\text{test}(x^i) = \exp\left(-\frac{(x^i - \mu^i)^2}{2\sigma^2}\right) \]

Genome-based Attack

\[ f_\text{test}(x^i) = \exp\left(-\frac{(x^i - \mu^i)^2}{2\sigma^2}\right) \]

Defense with Differential Privacy

\[ \text{D}: \text{methylation values of 60 patients} \]

\[ \text{D}': \text{one patient different} \]

\[ \text{output a random mean that hides the contribution of the changed entry} \]

\[ \Pr[M(\text{mean}(D)) = \mu] \leq \epsilon \Pr[M(\text{mean}(D')) = \mu] \]

where \( M(\text{mean}(D)) = \text{mean}(D) + \text{Lap}(\frac{\epsilon}{\text{number of positions}}, \text{number of patients}) \)

\[ \text{number of positions} = 300,000 \]

\[ \text{number of patients} = 60 \]

Abbreviation Description Tissue Type Number of Patients GEO identifier

| GBM | glioblastoma | brain cancer | 136 | GSE36270 |
| IN | prostatic adenocarcinoma | | 80 | GSE45484 |
| IBD CD | Crohn's disease | | 77 | GSE37460 |
| IBD UC | ulcerative colitis | | 79 | GSE37460 |
| BC | breast cancer | | 892 | |
| WGBS | genome and methylation data | | 75 | not publicly available |

inken.hagestedt@cispa.saarland
yang.zhang@cispa.saarland