Privacy in Epigenetics: Temporal Linkability of MicroRNA Expression Profiles

Michael Backes, *Saarland University and Max Planck Institute for Software Systems (MPI-SWS);* Pascal Berrang, Anna Hecksteden, Mathias Humbert, Andreas Keller, and Tim Meyer, *Saarland University*

https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/backes_epigenetics

This paper is included in the Proceedings of the 25th USENIX Security Symposium

August 10–12, 2016 • Austin, TX

ISBN 978-1-931971-32-4

Open access to the Proceedings of the 25th USENIX Security Symposium is sponsored by USENIX

Privacy in Epigenetics: Temporal Linkability of MicroRNA Expression Profiles

Michael Backes
backes@cispa.saarland
CISPA, Saarland University & MPI-SWS
Saarland Informatics Campus

Pascal Berrang
pascal.berrang@cispa.saarland
CISPA, Saarland University
Saarland Informatics Campus

Anne Hecksteden
a.hecksteden@mx.uni-saarland.de
Institute of Sports and Preventive Medicine
Saarland University

Mathias Humbert
mathias.humbert@cispa.saarland
CISPA, Saarland University
Saarland Informatics Campus

Andreas Keller
andreas.keller@ccb.uni-saarland.de
Clinical Bioinformatics, Saarland University
Saarland Informatics Campus

Tim Meyer
tim.meyer@mx.uni-saarland.de
Institute of Sports and Preventive Medicine
Saarland University

Abstract

The decreasing cost of molecular profiling tests, such as DNA sequencing, and the consequent increasing availability of biological data are revolutionizing medicine, but at the same time create novel privacy risks. The research community has already proposed a plethora of methods for protecting *genomic* data against these risks. However, the privacy risks stemming from *epigenetics*, which bridges the gap between the genome and our health characteristics, have been largely overlooked so far, even though epigenetic data such as microRNAs (miRNAs) are no less privacy sensitive. This lack of investigation is attributed to the common belief that the inherent temporal variability of miRNAs shields them from being tracked and linked over time.

In this paper, we show that, contrary to this belief, miRNA expression profiles can be successfully tracked over time, despite their variability. Specifically, we show that two blood-based miRNA expression profiles taken with a time difference of one week from the same person can be matched with a success rate of 90%. We furthermore observe that this success rate stays almost constant when the time difference is increased from one week to one year. In order to mitigate the linkability threat, we propose and thoroughly evaluate two countermeasures: (i) hiding a subset of disease-irrelevant miRNA expressions, and (ii) probabilistically sanitizing the miRNA expression profiles. Our experiments show that the second mechanism provides a better trade-off between privacy and disease-prediction accuracy.

1 Introduction

Since the first sequencing of the human genome in 2001, tens of thousands of genomes and over a million genotypes have been sequenced. The knowledge of our genetic background enables to better predict, and thus anticipate, the risk of developing several diseases, including cancers, cardiovascular and neurodegenerative diseases. Moreover, the genomic research progress enables the development of personalized treatment through pharmacogenomics, studying the effect of the genome on drug response. One of the most important negative counterparts of this genomic revolution is the threat towards genomic privacy [11, 39]. Genomic data contains very sensitive information about individuals' predisposition to certain severe diseases, about kinship, and about ethnicity, all of which can lead to various sorts of discrimination. Furthermore, genomic data is very stable over time and correlated between family members [28]. Therefore, a lot of research has already been carried out to improve the genomic-privacy situation (most of the related literature is surveyed in [20, 42]).

However, our genome is not the only element influencing our health. Environmental factors (e.g., pollution, diet, lifestyle, ...) often play a crucial role in the development of most common diseases. Epigenetics (or epigenomics), transcriptomics, and proteomics aim to bridge the gap between the genome and our health status. Multi-omics research is a logical complementary step to genome sequencing: the DNA sequence tells us what the cell could possibly do, while the epigenome and transcriptome tell what it is actually doing at a given point in time. Using a computer analogy, if the genome is the hardware, then the epigenome is the software [16].

Despite the growing importance of epigenetics in the biomedical community, privacy concerns stemming from epigenetic data have received little to no attention so far. With the increasing understanding of epigenetics, it becomes clear that epigenetic data contains a vast amount of additional sensitive information, and can thus raise potential privacy risks. For example, a large number of severe diseases (such as cancers, diabetes, or Alzheimer's [21, 33, 46, 53]) are already identified to be affected by epigenetic changes and a recent study found that epigenetic alterations could even affect sexual orientation [43]. Furthermore, epigenetic data can potentially

tell us more about whether someone is carrying a disease at a given point in time, compared to the genome that only informs about the *risk* of getting certain diseases.[1] Moreover, it is still unclear whether the current genetic nondiscrimination laws would apply to epigenetic data. For instance, the US Genetic Information Nondiscrimination Act (GINA) is limited to genetic characteristics and epigenetic data might not be considered genetic information [18, 47].

In this work, we focus on microRNAs (abbreviated miRNAs), an important element of the epigenome discovered in the early 1990s. MiRNAs are small RNA molecules that regulate the majority of human genes. Studies of miRNA expression profiles have shown that dysregulation of miRNA is linked to neurodegenerative diseases, heart diseases, diabetes and the majority of cancers [21, 33, 40, 46, 53].[2] Therefore, miRNA expression profiling is a very promising technique that could enable more accurate, earlier and minimally invasive diagnosis of major severe diseases. As a consequence, it will certainly be increasingly used in medical practice.

In contrast to the DNA sequence, which mostly stays constant over time, it is believed in the biomedical community that the miRNA expression levels are varying sufficiently to invalidate any linkability attempts over time, thus naturally protecting personal privacy. This work, however, shows the contrary: despite their temporal variability, microRNA expression profiles are still identifiable and linkable after time periods of several months.

Contributions. In this paper, we study the temporal linkability of personal miRNA expression profiles, by presenting and thoroughly evaluating different attacks, and proposing defense mechanisms to enhance unlinkability.

Specifically, we first study an identification attack, which pinpoints a specific miRNA expression profile in a database of multiple expression profiles by knowing the targeted profile at another point in time. Second, we study a matching attack, which tracks a set of miRNA expression profiles over time. We rely on principal component analysis to pre-process the miRNA expression levels, and on a minimum weight assignment algorithm for the matching attack. We thoroughly evaluate these linkability attacks by using three different longitudinal datasets: (i) the blood-based miRNA expression levels of athletes at two time points separated by one week, (ii) the plasma-based miRNA expression levels of the same athletes at two time points separated by one week, and (iii) the plasma-based miRNA expression levels of patients with lung cancer over more than 18 months and eight time points. Our experimental results show that blood miRNA expression profiles are about twice as easy to track over time compared to plasma miRNA profiles, and that the matching attack is more successful than the identification attack: We reach a success rate of 90% with blood and a success rate of 48% with plasma miRNAs in the matching attack whereas, in the identification attack, we reach a success rate of 76% with blood and 28% with plasma miRNAs. Moreover, we demonstrate that 10% of the miRNAs are already sufficient to achieve similar success rates as with all miRNAs. With the third dataset, we also observe that the attack achieves a similar success up to 12-month time periods.

We present two countermeasures to improve the unlinkability of miRNA expression profiles: (i) hiding a subset of the miRNA expressions, e.g., those that are not relevant for medical practice, and (ii) disclosing noisy miRNA expression profiles by adding noise in a differentially private and distributed manner. While the first countermeasure is useful especially in a clinical setting, in which the disease-relevant miRNAs are already known, the second countermeasure is intended to be better suited for the biomedical research community. In this context, as one of the objective is to discover associations between miRNAs and diseases, it is impossible to restrict the released data to only a few miRNAs.

We evaluate our protection mechanisms with the first aforementioned blood-based miRNA profiles of athletes and a fourth, also blood-based, miRNA dataset of more than 1,000 participants that includes information about 19 diseases (at a single point in time). The former is used to measure how temporal linkability is reduced with our countermeasures, whereas the latter helps us evaluate the evolution of accuracy (i.e., utility) in predicting patients' diseases from their miRNA expressions. The experiments show that it is possible to decrease linkability by at least 50% for almost no loss of accuracy ($< 1\%$) for the majority of diseases with the noise mechanism. Moreover, our results demonstrate that the noise mechanism provides better privacy-utility trade-offs than the hiding method in 17 out of 19 of diseases, while allowing more flexibility in the data usage for biomedical researchers. This finding is reinforced by the fact that an adversary could use correlations between miRNA expressions to infer more miRNA expressions than those actually shared by our first countermeasure.

Organization. In Section 2, we present the biomedical background relevant to understand our work. In Section 3, we introduce the adversarial model. We then describe in detail our four datasets in Section 4. In Section 5, we present the analytical tools used to carry out our linkability attacks and our experimental results. In Section 6, we propose and evaluate countermeasures and compare their performance. We present the related literature in Section 7 before concluding in Section 8.

[1] The only exception to this rule are Mendelian disorders, such as cystic fibrosis, which are largely determined by our genes.

[2] Known relations between miRNA and human pathologies can be found at `http://www.cuilab.cn/hmdd`.

2 Background

We briefly review the genetic concepts useful for understanding our paper. Epigenetics etymologically come from the combination of *epi*, which means "above", "over" in Ancient Greek, and *genetics*, which means "origin". This term broadly refers to the study of cellular and phenotypic trait variations stemming from other causes than changes in the genotype. These external factors are for example the in-utero or childhood development, environmental chemicals, aging or diet. Epigenetics can also refer to the changes themselves, such as DNA methylation and histone modification, which alter how genes are expressed without modifying the genome.

MicroRNAs (miRNAs) are epigenetically regulated mechanisms discovered in the early 1990s. MiRNAs are small non-coding RNA molecules that regulate gene expression in plants and animals. It has been shown that 60% of genes coding human proteins are regulated by miRNAs [25]. Whereas a miRNA is a RNA molecule containing around 22 nucleotides, *miRNA expression* is a real-valued number quantified in a two-step polymerase chain reaction (PCR) process. Different sets of miRNAs are expressed in different cell types and tissues.

Biomedical research is notably interested in discovering how miRNA expression affects physiological and pathological processes.[3] Studies of miRNA expression profiling have demonstrated that dysregulation of miRNA is linked to neurodegenerative diseases (Alzheimer's and Parkinson's), heart diseases, diabetes, and the majority of cancers [21, 33, 40, 46, 53]. MiRNA expression profiling is hence a very promising technique that could enable more accurate, earlier and minimally invasive diagnosis of severe diseases. To mention one current, concrete application, miRNA expressions taken from *blood* samples suffice to detect several diseases, such as cancer or Alzheimer's [34, 37]. In the following, we study the temporal linkability of miRNA expression profiles coming from blood and plasma (serum) samples.

3 Adversarial Model

We assume the adversary gets access to miRNA expression profiles of individuals at different points in time. Such epigenetic data is increasingly available in public research databases, such as the Gene Expression Omnibus (GEO) [4] or ArrayExpress [1] databases. Moreover, such data could be leaked through a major security breach, e.g., of a hospital server. Health data is also increasingly available on the black market. For instance, cyber attacks against healthcare companies have increased by 72% from 2013 to 2014 [3]. Moreover, 91% of healthcare companies have experienced a violation of their databases over the last two years, and only 32% feel they have adequate resources to defeat these incidents [6]. Real-world cyber attacks show us that health data can be hacked en masse [5, 8] or that attacks can be more targeted towards high-profile victims [9]. Very sensitive medical data of thousands of patients can also end up online due to a human mistake [2].

In a typical scenario, the adversary would get access to miRNA expression levels of one or multiple individuals from a (private) health insurance or hospital database, and wants to match them with a (public) research dataset of miRNA expression levels at another point in time. A particularly sensitive scenario would be the matching of non-anonymized healthy miRNA samples with miRNA profiles that are known to be associated with diseases. Also note that researchers have demonstrated that RNA expression profiles could be matched to genotypes by relying on expression quantitative trait loci (eQTLs) [48]. Therefore, if the adversary can also access the genotypes of the victims, these genotypes provide him with further means for de-anonymizing the corresponding (micro)RNA expression profiles [26, 30]).

4 Dataset Description

Unlike in other fields of privacy research, where large amounts of data can be collected in a small amount of time and at low cost, in the health-privacy field we face the exact opposite: measuring the miRNA expression levels of one single sample already costs several hundred dollars. Longitudinal epigenetic data are particularly valuable, since patients have to regularly provide their biological samples over a long period of time. Therefore, the four datasets used throughout the paper, and described hereunder, represent very rich data.

We start by describing our three longitudinal datasets. The first dataset contains the blood-based miRNA expression levels of 29 well-trained male athletes (15 endurance athletes and 14 strength athletes) at two points in time, while the second dataset contains the plasma-based miRNA expression levels of those athletes at the same points in time.[4] None of the athletes is known to be affected by a disease. The samples were taken prior and post exercising (time period of one week), similar to the data previously presented in [12]. The athletes followed a 6-day training with two training sessions a day, except at day 4 when only one session was scheduled.

[3] Strictly speaking, miRNA is part of the epigenome while miRNA expression is generally considered more as part of the transcriptome. In this paper, we use the term epigenetics in its broad acceptation.

[4] We selected blood and plasma since these two body fluids are likely candidates as source for biomarkers in future applications.

The tests were conducted at Saarland Unversity (Germany) for the endurance athletes, and at Ruhr University Bochum (Germany) for the strength athletes.

In order to confirm our results, we make use of a third, independent dataset. This dataset contains the miRNA expression data of plasma of 26 lung-cancer patients (9 females and 17 males) over a period of more than 18 months [38], at eight time points: before surgery (tumor resection), two weeks after surgery (abbreviated A.S. in the graphs), and 3, 6, 9, 12, 15, and 18 months after surgery.[5] The patients' ages range from 47 to 79. All three longitudinal datasets include the expression levels of 1,189 miRNAs for each individual at every time point.

Our last dataset contains the expression levels for 848 miRNAs collected from blood samples for each of 1,049 individuals [35] at only one time point. 94 of these individuals are considered to be healthy and are used as a control group in Section 6. Most of the rest represent cases, i.e., individuals carrying one out of the following 19 different diseases: 124 have Wilms tumor, 73 lung cancer, 65 prostate cancer, 62 myocardial infarction, 47 chronic obstructive pulmonary disease (COPD), 45 sarcoidosis, 45 ductal adenocarcinoma, 43 psoriasis, 37 pancreatitis, 35 benign prostate hyperplasia, 35 melanoma, 33 non-ischaemic systolic heart failure, 29 colon cancer, 24 ovarian cancer, 23 multiple sclerosis, 20 glioma, 20 renal cancer, 18 periodontitis, and 13 stomach tumor.

Note that a miRNA expression generally takes values between 0 (meaning the miRNA is not expressed at all) and tens of thousands. As we will mention later, we typically filter out miRNA whose median expressions among all individuals are smaller than 50, since these are non-expressed or not expressed enough to be significant.

While the last two datasets are both freely available in the GEO database (see accession number GSE68951 and GSE61741), the datasets consisting of athletes' miRNA expressions are not yet publicly available, but will be made available soon.[6] We also discuss ethical considerations and how we handled these datasets in Appendix A.1.

5 Linkability Attacks

We study the extent of the linkability threat (as described in Section 3) by means of two attacks. First, we describe the mathematical principles behind our attacks, and then evaluate their success on our three longitudinal datasets.

[5] Note that for the last two points in time, we have the miRNA profiles of 25 and 22 patients, respectively.

[6] Please contact Andreas Keller for more information about the access to these datasets.

5.1 The Attacks

The first attack, called *identification attack*, refers to a scenario in which the adversary knows the miRNA expression profile of a targeted individual and aims at finding the corresponding miRNA expression profile in a database of n miRNA expression profiles, e.g., later in time. The second attack, called *matching attack*, refers to the case where the attacker has access to two databases of miRNA profiles collected at different points in time and wants to match their elements together.

For both our attacks, as there are more than 1000 known miRNAs with real-valued expression levels, we apply a pre-processing step using principal component analysis (PCA) with whitening. In particular, we apply the probabilistic PCA model proposed by Tipping and Bishop [49], which relies on singular value decomposition. This PCA step projects the high-dimensionality miRNA expression vectors to smaller-dimensionality uncorrelated components. The whitening step divides the resulting PCA components by the number of samples multiplied by the singular values in order to provide uncorrelated expression vectors of unit variance. We then make use of the Euclidean distance between the miRNA expression vectors projected on the first c principal components.

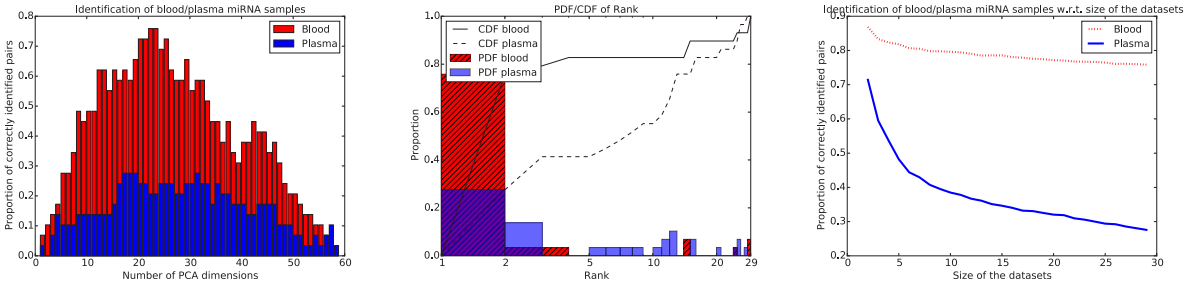
In the identification attack, we assume the adversary has had access to the miRNA profile $\mathbf{r}_k^{t_1}$, vector containing the miRNA expressions of an individual k at time t_1, and he wants to identify this individual in a database of n miRNA expression profiles $\{\mathbf{r}_i^{t_2}\}_{i=1}^n$ collected at time $t_2 \neq t_1$. After having extracted the c principal components from the whole dataset by using PCA, the adversary ranks the n profiles (projected on the c components) $\{\vec{\mathbf{r}}_i^{t_2}\}_{i=1}^n$ by decreasing distance to the targeted miRNA profile (also projected on the c components) $\vec{\mathbf{r}}_k^{t_1}$ and picks the profile with minimum distance to the targeted profile. Formally, the adversary will select the profile $\vec{\mathbf{r}}_{i^*}^{t_2}$ where

$$i^* = \arg\min_i \left\| \vec{\mathbf{r}}_i^{t_2} - \vec{\mathbf{r}}_k^{t_1} \right\|_2 .$$

In the matching attack, the adversary has access to two databases of miRNA expression profiles at two different time points t_1 and t_2. We assume that the databases are of sizes n_1 and n_2, both strictly greater than 1. First, if $n_1 = n_2 = n$, the adversary will assign one miRNA profile at time t_1 to exactly one profile at time t_2. In this case, the best assignment σ^* is the one that minimizes the sum of the distances between every matched pair:

$$\sigma^* = \arg\min_\sigma \sum_{i=1}^{n} \left\| \vec{\mathbf{r}}_{\sigma(i)}^{t_2} - \vec{\mathbf{r}}_i^{t_1} \right\|_2 .$$

This problem boils down to finding a perfect matching on a weighted bipartite graph, with n vertices on both



(a) (b) (c)

Figure 1: Success rate of the identification attack for the athletes dataset. (a) Proportion of successfully identified pairs plotted against the number of PCA dimensions (in $\{1,...,58\}$). (b) Probability density function (PDF) and cumulative distribution function (CDF) of obtained ranks. (c) Proportion of successfully identified pairs plotted against the number of miRNA expression profiles.

sides representing the miRNA profiles, and a weight on each edge representing the Euclidean distance between any pair of miRNA profiles (vertices), projected on the first c principal components. We want to find the matching among $n!$ possible assignments that minimizes the sum of the weights between vertices. Fortunately, there exist several algorithms in the literature that find the minimum weight assignment in polynomial time. We use the blossom algorithm [19], because it only has a complexity of $O(n^3)$ and it can also be applied to general graphs.

If $n_1 \neq n_2$, we fill the smallest side of the bipartite graph with dummy vertices. Then we assign infinite weight to all edges from actual vertices to these dummy vertices in order to ensure that the dummy vertices will be the least likely assigned to the vertices in the largest side that are also present in the smallest side.

5.2 Experimental Results

We evaluate how successful both aforementioned attacks are in breaking the privacy of our three longitudinal datasets. We implement the attacks in Python, and make use of the libraries Scikit-learn [13, 44] (for PCA) and NetworkX[7] (for the graph matching).

5.2.1 Identification Attack

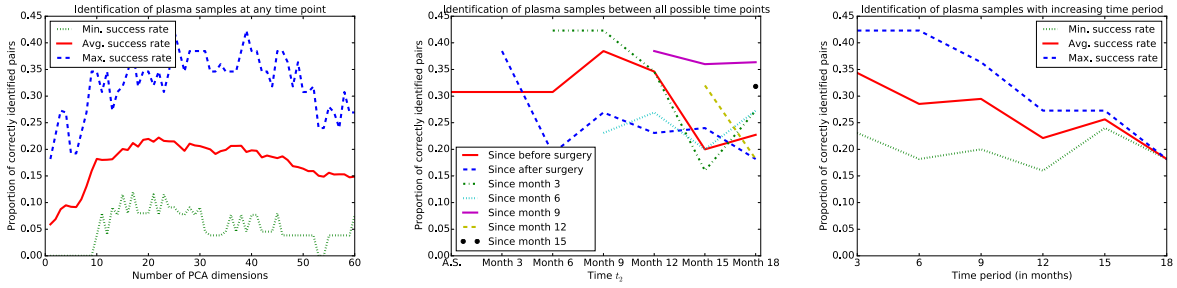
In this subsection, we evaluate the success of an adversary, who aims at identifying the miRNA profile of a targeted individual in a longitudinal dataset. As mentioned in Section 4, the first two longitudinal datasets contain miRNA expression levels of 29 individuals collected at a time interval of one week.

First, we compare the success rate for correctly identifying samples for all possible PCA dimensions. Fig. 1(a) indicates that the blood's miRNA expression levels are easier to identify over time than the plasma's miRNA expression levels. When identifying samples by their blood miRNA expression levels, we can reach a maximum success rate of 76% for the blood with 22 or 23 PCA dimensions. The maximum success rate for the plasma is 28% with 17, 18, 19 or 31 PCA dimensions. Note that both achieve their highest success with a number of PCA dimensions around 20.

Next, we rank the miRNA profiles at time t_2 in order of increasing distance to the targeted profile $\mathbf{r}_k^{t_1}$. Fig. 1(b) shows the rank of the correct sample $\mathbf{r}_k^{t_2}$ by using 22 PCA dimensions for the blood and 18 PCA dimensions for the plasma. The correct profile is ranked within the top 2 profiles in more than 40% of the cases for the plasma, whereas the correct sample is ranked withing the top 2 samples in 80% of the cases for the blood.

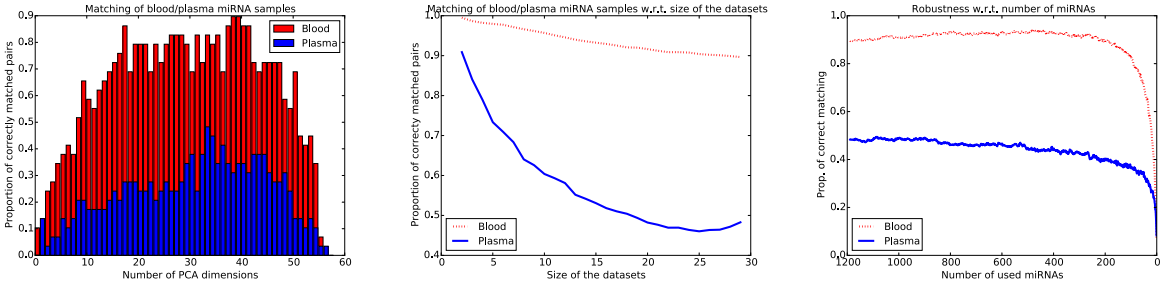
In order to get an impression on the attack's performance on larger datasets, we also analyze the success of the identification attack with respect to the number of participants in the dataset, i.e., we vary the number of profiles among which the attacker has to identify the targeted miRNA profile, again using 22 PCA dimensions for the blood and 18 PCA dimensions for the plasma. Intuitively, when the number of miRNA samples increases, the success rate of the attacker should decrease. In this experiment, we adjust the number n of miRNA profiles between 2 and 29 and evaluate the attacker's success on a subset of our datasets. In particular, for each number of profiles n, we randomly choose 1000 different combinations (or fewer if necessary) of n out of 29 miRNA profiles and run the identification attack on every sample within this subset. Fig. 1(c) depicts the average success rates for each number of profiles n. As expected, the success rate monotonically decreases with the number of participants for blood and plasma samples. For plasma,

[7] `https://networkx.github.io`



(a) (b) (c)

Figure 2: Success rate of the identification attack for the lung cancer dataset. (a) Success rate aggregated over all identifications between any t_1 and t_2 plotted against the number of PCA dimensions. (b) Success rate of identifying the miRNA profiles between time pairs t_1 and t_2. (c) Success rate plotted against the time period between t_1 and t_2.



(a) (b) (c)

Figure 3: Success rate of the matching attack for the athletes dataset. (a) Proportion of successfully matched pairs plotted against the number of PCA dimensions. (b) Proportion of successfully matched pairs plotted against the number of miRNA profiles. (c) Proportion of successfully matched pairs plotted against the number of revealed miRNAs.

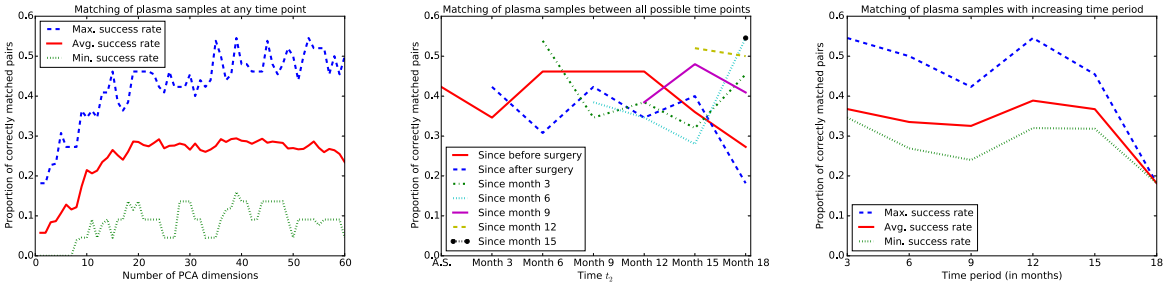
however, this decrease is much sharper, confirming that the blood's miRNA expression levels provide means for easier identification. From the curves' slopes, we can predict that, for larger datasets, blood based samples will still be subject to a relatively high identification success.

In order to validate our findings, we also evaluate our experiments on our other longitudinal, independent dataset containing plasma miRNA profiles from 26 individuals with lung cancer collected over up to eight different points in time.

First, we evaluate the attacker's success with respect to a varying number of PCA dimensions. Fig. 2(a) depicts the minimum, average and maximum success rate of an attacker when identifying the samples between different points in time, irrespective of the time period between them. The maximum success rate for the identification attack is 42% and is achieved for 25 and 39 PCA dimensions. The usage of 22 PCA dimensions yields the highest average success rate, of 22%. The highest minimal success rate in the dataset is achieved for 17 PCA dimensions (12%).

These results are similar to what we obtained in our experiments for the athletes dataset: The best results are achieved for a number of PCA dimensions around 20 in both datasets. The highest average success rate lies 6 points below the best success rate for the athletes dataset. This could be explained by longer time periods in this dataset. However, for some time periods, we can achieve one and a half the success rate of the first dataset. When comparing the top 10 miRNAs contributing to the first PCA dimension in this dataset and in the athletes' plasma dataset, we also find an overlap of 80% between these miRNAs. This indicates that approximately the same set of miRNAs can be used to differentiate plasma expression profiles between individuals in both datasets. Thus, we can conclude that, while miRNA expression levels are directly linked to health status, the health status only affects a subset of the miRNAs, which has only little effect on the temporal linking.

To further investigate the effect of different time periods on the attacker's success, we plot the maximum success rates between all possible, ascending combinations



(a) (b) (c)

Figure 4: Success rate of the matching attack for the lung cancer dataset. (a) Success rate aggregated over all matchings between any t_1 and t_2 plotted against the number of PCA dimensions (in $\{1,...,60\}$). (b) Success rate of matching the miRNA profiles between time pairs t_1 (various curves), t_2 (x-axis value). (c) Success rate plotted against the distance between t_1 and t_2.

of points in time in Fig. 2(b). With only a few exceptions, the best success rates are most often achieved for consecutive time points. The only two exceptions are found for t_1=*before the surgery* and t_1=*the sixth month after the surgery*. In general, however, we notice a tendency of slight decrease in success over an increasing time period.

In order to verify this finding, we group the results by the period between t_1 and t_2 (Fig. 2(c)). Note that, since we do not know the time period between before the surgery and after it, we leave out all results that use samples collected before the surgery. Clearly, the best achievable success rate drops for increasing time periods.This decrease over larger periods of time can partially explain the lower average success rate in this dataset compared to the athletes' dataset (considering a much smaller time period).

Next, we computed the guessing entropy [14, 41] for the identification attack. The guessing entropy $\mathrm{E}[G(X)]$ is the expected number of guesses an adversary would need to identify the correct sample at a different point in time. For the identification attack it is given by $\mathrm{E}[G(X)] = \sum_{i=1}^{n} i \cdot \Pr[X = i]$, where X denotes a the rank of the correct sample at time t_2 and $\Pr[X = i]$ denotes the empirical probability that the correct sample is ranked at the i^{th} position.

For blood-based samples of our athletes dataset, the attack can achieve a guessing entropy just below 4, clearly outperforming random guesses, which would yield an entropy of 15 guesses on average. For plasma-based samples of the same dataset, the attack yields an entropy of approximately 9 guesses. This result is consistent with the results on the lung cancer dataset, where, on average, an adversary would need just fewer than 9 guesses (compared to a guessing entropy of 13.5 for random guesses). Moreover, for some t_1 and t_2, the attack is even able to achieve a guessing entropy smaller than 6.

5.2.2 Matching Attack

We evaluate here the success of the adversary, who tries to link all participants over time, again for the three aforementioned longitudinal datasets. Starting with the athletes' datasets, we compare the success rate of matching the blood and the plasma over all possible PCA dimensions for 29 participants. In Fig. 3(a), we notice the same behavior as in the identification attack: the blood based miRNA expression levels are much easier to link over time than the plasma based levels. We even reach a higher maximum absolute success rate than in the identification attack: 90% with 39 or 40 PCA dimensions for the blood and 48% success with 34 PCA dimensions for the plasma samples.

The identification attack's lower success rate is due to the fact that it is evaluated for each sample individually, thus allowing multiple samples at t_1 to be linked to the same (potentially wrong) sample at t_2. Since our perfect matching attack rules out those cases by forcing each profile at t_2 to be matched to exactly one profile from t_1, it also decreases the number of wrongly matched samples.

Next, we also analyze the success of the attack with respect to the number of participants to be matched together. Intuitively, the more miRNA profiles there are, the more challenging it should be for the adversary to match them at different time points. Again, we make the number of participants n vary between 2 and 29 at both time points, again randomly sampling 1000 combinations (or fewer, if there are fewer than 1000 combinations) and averaging the result. Fig. 3(b) shows the expected trend of decreasing success for the blood miRNA samples. The plasma scenario monotonically decreases between 2 and 25 participants and then slightly increases until 29. This artifact could be explained by the smaller number of random combinations, and thus experiments, when $n > 26$. We also find that the blood attack faces

a rather linear decrease in success whereas the plasma success rate decreases much faster. By extrapolating this linear trend, we can expect a success rate as high as 60% with 120 participants in the datasets. Therefore, we again conclude that the blood has miRNA expression levels that enable much easier tracking over time than the plasma, which is consistent with the results of the identification attack.

Fig. 3(c) investigates how the attack's success evolves when revealing only a subset of the miRNA expression levels. We gradually drop individual miRNAs in random order and compute the attack's success. The figure shows the success rate (for each possible number $m \in \{1189, 1188, \ldots, 2, 1\}$ of miRNAs) averaged over 50 randomly chosen orderings of miRNAs. We notice that the attack's success is very stable, especially for the blood samples, from 1189 to 200 miRNAs. For the blood, the success decreases below 80% the first time when there are fewer than 100 miRNAs available to the adversary. We further study the implications of this robustness in the context of our countermeasures in Section 6.

We also made use of our third longitudinal dataset containing plasma miRNA expression profiles of 26 individuals over up to eight different time points (cf. Section 4). In Fig. 4(a), we see that the average success rate reaches its maximum at a number of PCA dimensions very close to the number of dimensions for the athletes dataset, i.e., 34. However, this maximum is approximately 30%, which is smaller than the 48% reached for the first dataset. A greater period between time points could explain this behavior, and we also see that we can still reach a maximum success rate of 55% between some time points, with 39 PCA dimensions. We explore the time effect in deeper details in the following figures.

Fig. 4(b) depicts the maximum success rate between any pair of time points t_1, t_2. For instance, the solid red line shows the success rates between t_1=*before surgery* and all others. It is difficult to detect any trend with respect to the time period in the different curves, except a slight decrease when the time period is higher or equal to 15 months. This is confirmed by Fig. 4(c) that depicts the maximum, average, and minimum success rate with respect to the period between t_1 and t_2. We clearly notice a decreasing rate between 3 and 9 months, an increase to 12 months, and finally clear decrease towards 15 and 18 months.

6 Countermeasures

In this section, we propose and evaluate two main defense mechanisms for preventing miRNA expression data from being tracked over time. The proposed techniques are based on well-established privacy-enhancing methods, previously applied in other privacy contexts, such as location privacy. The first approach relies on a quite straightforward technique: release only a subset of the miRNAs. We can already see from Fig. 3(c) of Section 5 that the matching attack is quite robust to a decrease in the number of miRNAs. Nevertheless, we show hereafter how we can keep a high utility in combination with unlinkability of expression profiles over time by revealing a small subset of miRNAs. The second countermeasure consists in adding noise to the released miRNA expression vectors, independently for every individual. This method shows very promising results, reaching an even better privacy-utility trade-off than the hiding mechanism. Furthermore, we also investigate the effect of correlations between miRNA expression levels and present the privacy evolution when the adversary can infer missing miRNAs by using these correlations.

For evaluating the privacy provided by our defense mechanisms, we focus on the matching attack against blood-based miRNAs, as this constitutes the worst-case attack from a privacy perspective, as shown in Section 5. Moreover, we assume the attacker is able to select the number of PCA dimensions that maximizes his success. This provides us with a conservative measure of privacy, showing the worst-case privacy levels individuals can expect.

6.1 Baseline Utility

Before presenting the proposed countermeasures and their efficiencies, we must carefully define the context in which they should apply. Indeed, we can rarely have both perfect privacy and maximum utility, so that we often need a trade-off between these two. Therefore, the efficiency of the defense mechanism cannot only be judged based on the privacy metric, but must also relate to the utility brought in the context in which the data is used.

According to biomedical experts, miRNA expression profiles have strong potential to help predict various severe diseases, from cancer to Alzheimer's disease. Biomedical researchers typically rely on standard machine learning algorithms to identify which miRNAs are playing a significant role in the disease of interest. They are dealing with binary classification, between cases (carrying the disease) and controls (healthy), and most often rely on support vector machines (SVMs). In particular, they typically use radial basis function SVMs and select a subset of features by subsequently adding miRNAs in order of their significance values (e.g., p-values computed by the Wilcoxon-Mann-Whitney (WMW) test) [37] or equivalently in order of their area under the ROC curve (AUC). Given samples of cases and controls, the accuracy is then defined as the number of correctly classified samples divided by the

Disease	Maximum accuracy with the best subset of expressed miRNAs (# miRNAs)	Accuracy with all expressed miRNAs
Periodontitis	0.941 (37)	0.88
Renal cancer	0.988 (32)	0.962
Wilms' tumor	0.95 (150)	0.937
Benign prostate hyperplasia	0.921 (105)	0.883
Chronic obstructive pulmonary disease	0.932 (70)	0.886
Colon cancer	1.0 (30)	0.997
Ductal carcinoma	0.938 (55)	0.92
Glioma	0.927 (19)	0.83
Lung cancer	0.899 (60)	0.848
Melanoma	0.996 (185)	0.992
Multiple sclerosis	0.992 (40)	0.979
Myocardial infarction	0.893 (400)	0.884
Nonischaemic systolic heart failure	0.9 (135)	0.871
Ovarian cancer	0.919 (18)	0.876
Pancreatitis	0.941 (130)	0.899
Prostate cancer	0.923 (90)	0.91
Psoriasis	0.914 (350)	0.902
Sarcoidosis	0.977 (200)	0.97
Tumor of stomach	0.969 (160)	0.89

Table 1: Accuracy of the SVM algorithm in classifying individuals between cases (carrying the disease) and controls (healthy), for 19 diseases, without countermeasure.

total number of samples. Note that we compute the average accuracy over a repeated k-fold cross-validation.

In this work, we define the utility as the accuracy of the SVM classifier, as defined above. We use a 10-fold cross-validation with 5 repeats (using R and the caret[8] library) and determine the miRNAs' p-values by using the WMW test and adjusting the significance values for multiple tests using the Benjamini-Hochberg adjustment. The WMW test statistic is applied for each miRNA individually in order to test whether this miRNA has similar expressions between cases and controls (null hypothesis). The p-values then provide us with the relevance of the miRNA to the disease of interest. In contrast to the t-test, the WMW test can be applied on unknown distributions. This way, we follow the standard procedure of biomedical research. Table 1 shows the accuracy of our SVM algorithm applied on our 1000+ participants dataset to predict 19 diseases, without any obfuscation. The maximum accuracy here is what we refer to as the baseline utility in the subsequent subsections.

Note that, before running the SVM algorithm, we filter out non-expressed miRNAs, i.e., those with a median level of expression smaller than 50 over the 1000+ individuals, which leaves us with 446 expressed miRNAs.

6.2 Hiding MicroRNA Expressions

The first countermeasure that we study is miRNA expression hiding. This obfuscation technique has the advantage to be non-pertubative, i.e., to preserve the correct values of all revealed miRNA expressions. However, as we have seen in Section 5, the attacks are extremely robust to removal of miRNAs. In the following, we want to find an optimal trade-off between the diagnosis accuracy, i.e., the utility, and the unlinkability of the data, i.e., the privacy. To this end, we make use of both our blood-based datasets, the 1000+ dataset with blood-based miRNA expressions to run our SVM algorithm and the athletes' dataset with blood-based miRNAs to evaluate the level of privacy. Note that we filter both datasets' miRNAs in order to have the same set of 446 miRNAs in both cases. While we measure the utility in terms of accuracy of the SVM, the privacy is measured in terms of the maximum achievable success rate (over all possible PCA dimensions) of our matching attack.

Figure 5 shows the evolution of privacy and utility for a range of 1 to 100 disclosed miRNAs, for 6 different severe diseases.[9] We focus on this range of miRNAs as: (i) for more than 100 miRNAs, the attack's success rate is approximately the same as the one without countermeasure, and (ii) the SVM can already achieve very high accuracy with up to 100 miRNAs. We gradually reveal the miRNAs in decreasing order of significance (based on p-values), as computed in Subsection 6.1.

Figure 5 demonstrates that there exists a trade-off between the utility of miRNA expressions and the privacy of the contributors' data. Note that we also depict the relative decrease in accuracy compared to the maximum SVM accuracy computed in Subsection 6.1 and the relative decrease in the attack's success (increase in privacy) compared to the attack's success with all miRNAs, i.e., 90%. We see that the relative decrease in accuracy is almost always smaller than 10%. The only exceptions to this are with pancreatitis and melanoma, for fewer than 3 disclosed miRNAs. Moreover, regarding the privacy, the figures show that we can never reduce the attack's success by more than 50% when revealing more than 20

[8] `caret.r-forge.r-project.org`

[9] These are representative of the behavior of all 19 diseases we tested our privacy-preserving mechanisms on.

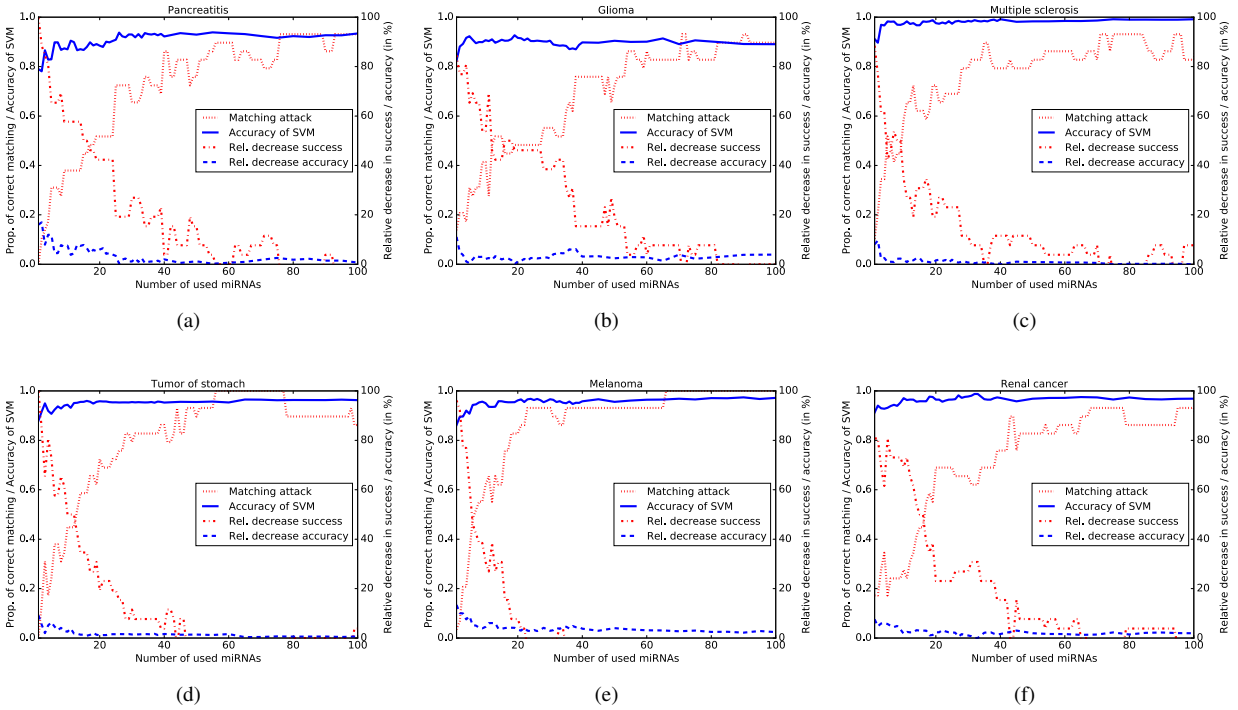
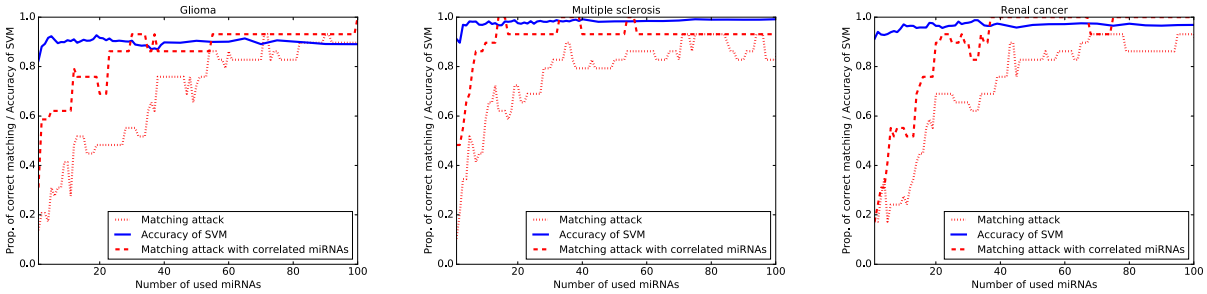


Figure 5: Evolution of privacy (unlinkability) and utility (classifier accuracy) plotted against the number of released miRNAs for the following diseases: (a) Pancreatitis, (b) Glioma, (c) Multiple sclerosis, (d) Tumor of stomach, (e) Melanoma, (f) Renal cancer. The *relative decrease success* curve refers to the decrease in success of the matching attack compared to the success without countermeasure. Similarly, the *relative decrease accuracy* curve refers to the decrease in accuracy of the SVM classifier with respect to the case without protection mechanism.

miRNAs. Nevertheless, within the range of 3 to 20 disclosed miRNAs, we can find, for all diseases, a satisfactory trade-off between utility and privacy.

In particular, for glioma, we can decrease the linkability attack's success and thus improve the privacy by 80.8% when using 4 miRNAs, while reducing the classification accuracy by only 1.1%. Similarly for multiple sclerosis, 7 miRNAs provide an increase in privacy of 53.8%, while the decrease in accuracy only amounts to 0.9%. For renal cancer and 10 miRNAs, we are able to achieve an improvement in privacy of 69.2% and a decrease of accuracy of only 1.7%. There are only two diseases for which it is very difficult to have both unlinkability and very high utility: melanoma and pancreatitis. For melanoma, we notice that the matching attack's success has a fast increase with very few miRNAs, and already exceeds 50% starting with only 7 miRNAs. For pancreatitis, the SVM's accuracy is relatively low (compared to the maximum) for the first 20 miRNAs. Thus for both diseases, either privacy or utility would have to be slightly sacrificed for the other.

MiRNA co-expression. Like between variants in the genome, there exist correlations between miRNA expressions: around 40% of miRNAs are not independently expressed [7]. This means that the adversary, by knowing these correlations, could increase his knowledge about the non-disclosed miRNA expressions. In order to evaluate the importance of such correlations, we first compute the Pearson's correlation coefficients, and their corresponding p-values, in all 99,235 pairs of the 446 expressed miRNAs in our fourth dataset. Filtering out all correlations with p-values greater than 0.001 (after Bonferroni correction for multiple correlations' testing) or correlation coefficient smaller than 0.5 leaves us with 47% of miRNAs not independently expressed. Figure 6 shows the updates of the linkability attack's success by taking into account all significant correlations as defined above. In our experiments, we take a quite conservative approach: We assume that the adversary can perfectly infer the miRNAs correlated with those that are gradually disclosed. The dotted curve provides an upper bound estimate on the success rate. A tighter bound could be derived by knowing more precisely the probabilistic dependencies between miRNAs. This is left for future work.



(a) (b) (c)

Figure 6: Correlations between miRNAs. Evolution of privacy and utility, when miRNAs correlated with the revealed miRNAs are taken into account for the attack. This provides an upper bound on the best linkability of miRNA expression profiles, i.e., worst-case privacy level. (a) Glioma, (b) Multiple sclerosis, (c) Renal cancer.

For Fig. 6, we make use of the three diseases of Figure 5 that gave best trade-off between privacy and utility, i.e., glioma, multiple sclerosis and renal cancer. We observe that the success rate knowing miRNAs correlated with disclosed miRNAs is much higher than without them, except for the very first miRNAs in Fig. 6(c). It shows that the most significant miRNAs for the SVM classification are co-expressed with others, which penalizes privacy significantly. Making use of the best subsets of miRNAs found above without correlations, containing 4 miRNAs for glioma, 7 for multiple sclerosis, and 10 for renal cancer, we evaluate the new privacy levels when miRNA correlations are taken into account. For glioma, instead of improving unlinkability by 80.8%, the 4 miRNAs and their correlated miRNAs yields an improvement in privacy of 34.6%. For renal cancers, the privacy enhancement drops from 69.2% to 38.5% and, for multiple sclerosis, knowing 7 miRNAs and their co-expressed miRNAs yield an attack's success rate almost equal to the highest rate with the full set of miRNAs. However, we can find new, better trade-offs: e.g., disclosing 5 miRNAs for multiple sclerosis still provides the same high SVM accuracy (decrease of 0.9% compared to the baseline) while reducing the attack's success by 23%. Note that we do not make use of the correlated miRNAs for the SVM algorithm as we are not certain about how they correlate with the disclosed ones.

6.3 Noise Mechanism

As we have noticed in the first protection mechanism, it is possible to hide the vast majority of miRNAs while retaining a fair level of prediction accuracy. This is typically very useful in the clinical setting where medical practitioners already know the miRNAs to test for predicting a specific disease. However, such a privacy-preserving mechanism could dramatically jeopardize miRNA utility for biomedical research. Indeed, as we have seen in our previous experiments, the majority of miRNAs need to be masked in order to gain a significant amount of unlinkability, which is not possible if researchers want to test for associations between miRNAs and diseases. Therefore, we additionally present and study a countermeasure where contributors of miRNA expressions directly apply random noise to their vectors of expression levels before providing them to the research community (possibly online), in a fully distributed manner (i.e., independently of other contributors).

The idea behind adding noise to the raw expression data is to provide indistinguishability between different expression vectors and consequently reduce the tracking capabilities of the adversary. Following the generalized notion of differential privacy [15] previously applied to location privacy [10], we state that a mechanism K achieves *epigeno-indistinguishability* if and only if, for all m-miRNA expression vectors $\mathbf{r}_1$, $\mathbf{r}_2$,

$$Pr(K(\mathbf{r}_1) \in \mathscr{S}) \leq exp(\varepsilon d_2(\mathbf{r}_1, \mathbf{r}_2)) \times Pr(K(\mathbf{r}_2) \in \mathscr{S}),$$

where $\mathscr{S}$ is any subset of the set of possible responses and $d_2(\cdot,\cdot)$ denotes the Euclidean distance. In the following, we assume the set of possible responses lies in the same m-dimensional real-valued space $\mathbb{R}^m$ as the set of original expression vectors. Before defining our mechanism $K(\cdot)$ for achieving epigeno-indistinguishability, let us first give some intuition about the mechanism. The noise mechanism is such that the probability of reporting a noisy expression vector $K(\mathbf{r})$ differs by at most a factor $exp(\varepsilon d_2(\mathbf{r}_1, \mathbf{r}_2))$ when the actual, non-obfuscated miRNA expression vectors are $\mathbf{r}_1$ and $\mathbf{r}_2$. This can be achieved by relying on the multivariate Laplacian mechanism that adds noise $\mathbf{x}$ according to the following probability density function $g(\mathbf{x}) = \frac{1}{\alpha} e^{-\varepsilon \|\mathbf{x}\|_2}$, where α is a normalization factor ensuring that the integral over all $\mathbf{x} \in \mathbb{R}^m$ equals one.

Sampling noise from the distribution $g(\mathbf{x})$ can be carried out efficiently by generalizing the method used for the planar Laplacian mechanism in [10]. First, we sample the magnitude $\|\mathbf{x}\|_2$ of the noise from a gamma distribution with shape m and scale $1/\varepsilon$. Second, we randomly generate the direction $\hat{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|_2$ of the noise by uniformly sampling points on the surface $\mathbb{S}^{m-1}$ of a hypersphere [36]. To do so, we can generate m independent Gaussian random variables y_1, y_2, ..., y_m, and let $\hat{y}_i = y_i/\sqrt{y_1^2 + \ldots + y_m^2}$ for $i = 1, ..., m$. Then the distribution of the vector $\hat{\mathbf{y}} = (\hat{y}_1, ..., \hat{y}_m)$ is uniform over the surface $\mathbb{S}^{m-1}$, and thus we can set the direction $\hat{\mathbf{x}} := \hat{\mathbf{y}}$. Each person i contributing his miRNA expression profile $\mathbf{r}_i$ will then share, instead of the actual expression data, the noisy vector $K(\mathbf{r}_i) = \mathbf{r}_i + \mathbf{x}$, where $\mathbf{x}$ is independently generated for all participants $i = 1, ..., n$.

Following this approach, in our evaluation, we first add noise to our dataset of 1000+ individuals (considering only the 446 miRNAs as before). Then, in the second step, we calculate the p-values on the noised data (since the researchers would be provided with exactly this data) and train the SVM the same way as in the previous subsections by subsequently adding miRNAs in the order of their p-values. Similarly, we evaluate the success of our attack on the athletes' dataset, when considering the same 446 miRNAs, but after adding noise. Moreover, we repeat both our experiments 50 times and average the results over all runs.

Figure 7 shows the evolution of the SVM accuracy and linkability (success of the attack), with respect to the amount of noise, tuned by ε, that is added to each contributor's miRNA expression profile. As privacy is measured on the same dataset for all six figures, its evolves in a very similar way. Even if the noise is randomly generated, the differences average out with the Monte Carlo method we use. We clearly see that with $\varepsilon = 1$, there is almost no privacy gain compared to the attack without countermeasure, whereas for $\varepsilon = 0.001$, the attack's success drops by almost 90%. Of course, as for the first countermeasure, there is a utility-privacy trade-off to be found between these two extreme values.

In Figure 7(a), we can observe that, for pancreatitis, $\varepsilon = 0.075$ is a good trade-off, with an accuracy decrease of only 0.8% and an unlinkability improvement of 40%. For glioma (Figure 7(b)), the best trade-off is certainly at $\varepsilon = 0.05$, with an accuracy decrease of 1.2% and an unlinkability improvement of 51%. For multiple sclerosis, we reach the best trade-off at $\varepsilon = 0.025$ with an accuracy decrease of 0.65% and an unlinkability improvement of 63%. For tumor of stomach, we can reach an accuracy decrease of only 0.2% and still improve the unlinkability by as much as 70% with $\varepsilon = 0.01$. For renal cancer, we have to sacrifice a bit more of utility, 2.3%, for a privacy increase of 61%, with $\varepsilon = 0.025$. The only disease for which it is quite difficult to get both satisfactory unlinkability and excellent accuracy is melanoma (Figure 7(e)). This is consistent with the hiding mechanism presented in Subsection 6.2, where we observed (in Figure 5(e)) a fast and sharp increase in the linkability attack's success.

6.4 Comparison of Protection Mechanisms

In order to compare both approaches, we decide upon a utility or a privacy requirement, fix it, and then evaluate the best privacy, respectively utility, achieved with both countermeasures. We carry out this evaluation on all 19 diseases for different requirements of utility and privacy.

First, we start by fixing the utility, more precisely the relative accuracy decrease compared to the baseline accuracy. The privacy is measured in terms of the decrease in the matching attack's success. For a given maximal decrease in accuracy $\Delta_{\mathsf{acc}}^{\mathsf{max}}$, we select the optimal number of miRNAs m^* and the optimal amount of noise ε^* that maximize the privacy increase $\Delta_{\mathsf{priv}}^{m}$ and $\Delta_{\mathsf{priv}}^{\varepsilon}$. In case of the hiding mechanism, we select $m^* = \arg\max_m \Delta_{\mathsf{priv}}^{m}$ such that $\Delta_{\mathsf{acc}}^{m} < \Delta_{\mathsf{acc}}^{\mathsf{max}}$. In case of the noise mechanism, we select $\varepsilon^* = \arg\max_\varepsilon \Delta_{\mathsf{priv}}^{\varepsilon}$ such that $\Delta_{\mathsf{acc}}^{\varepsilon} < \Delta_{\mathsf{acc}}^{\mathsf{max}}$, respectively.

Considering $\Delta_{\mathsf{acc}}^{\mathsf{max}} \in \{0.5\%, 1\%, 2\%, 3\%, 4\%, 5\%\}$ for all 19 diseases, we mostly experience that the noise mechanism provides a better privacy improvement compared to hiding a subset of miRNAs (all results are in Table 2 in the appendix). In particular, 90 out of 114 cases (combinations of disease and $\Delta_{\mathsf{acc}}^{\mathsf{max}}$) yield a better privacy with the noise mechanism. When examining a maximal decrease in accuracy of 2%, the hiding technique provides a better privacy for only 2 diseases, namely glioma and renal cancer. Interestingly, these two diseases stand out also for other values of the maximal accuracy decrease, providing better privacy with the hiding technique in 10 out of 12 cases. However, for all other diseases, adding noise in a distributed manner to individual expression profiles provides better utility for similar levels of privacy. For example, for lung cancer, we are able to achieve an increase in privacy of 79.3% while maintaining a decrease in accuracy of 0.8% using noise with $\varepsilon = 0.005$. The best we can achieve for the hiding technique here is either a decrease in accuracy of 0.97% and an increase in privacy of only 46.2% or a larger decrease in accuracy of 1.9% and a privacy improvement of only 50%.

Next, we discuss the results for a fixed minimal improvement of the privacy and compare the corresponding minimal decrease in accuracy in both countermeasures. We now fix the minimal increase in privacy (i.e., the minimal decrease in the attack's success) $\Delta_{\mathsf{priv}}^{\mathsf{min}}$ and minimize the decrease in accuracy: $\arg\min_m \Delta_{\mathsf{acc}}^{m}$ such

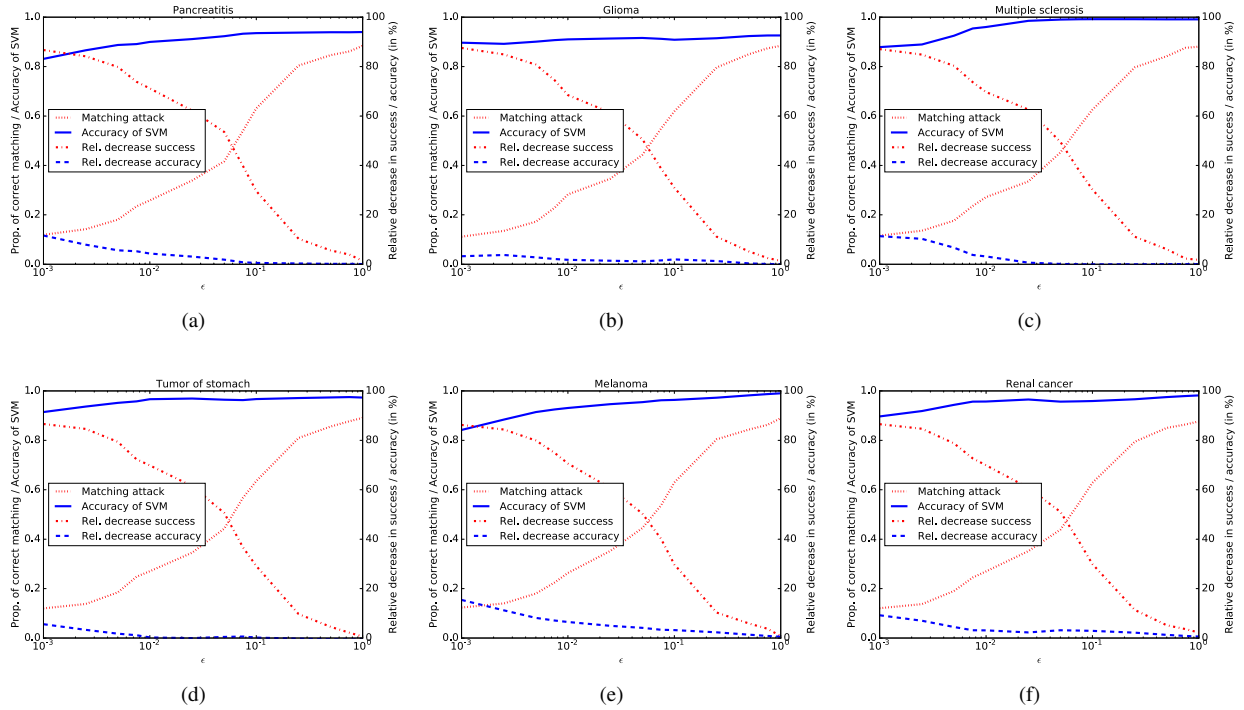


Figure 7: Evolution of privacy and utility (classifier accuracy) plotted against the noise (tuned by ε) added to the individual miRNA expression profiles, for the following diseases: (a) Pancreatitis, (b) Glioma, (c) Multiple sclerosis, (d) Tumor of stomach, (e) Melanoma, (f) Renal cancer.

that $\Delta^{m}_{\text{priv}} > \Delta^{\min}_{\text{priv}}$ and $\arg\min_{\varepsilon} \Delta^{\varepsilon}_{\text{acc}}$ such that $\Delta^{\varepsilon}_{\text{priv}} > \Delta^{\min}_{\text{priv}}$, respectively. We run experiments for values of $\Delta^{\min}_{\text{priv}}$ from 10% up to 90%, in steps of 10% (all results are provided in Table 3 in the appendix).

We again observe that, for most of the evaluated cases, the achieved accuracy is better when adding noise than when hiding miRNAs. In particular, this holds true for 143 out of 171 cases, clear exceptions being again glioma and renal cancer. For those two diseases, the hiding technique provides better accuracy than the noise mechanism in 87.5% of the cases. When fixing the minimal increase in privacy to 70%, only these two diseases provide better results with the hiding technique. For instance, with renal cancer, we achieve 60.8% improvement in privacy with a decrease in accuracy of 2.3% using noise with $\varepsilon = 0.025$, whereas we can obtain an increase in privacy of 69.2% and a decrease in accuracy of only 1.7% when using the hiding technique. For the majority of diseases, however, it is clearly the noise mechanism that provides much higher utility. For example, for lung cancer, an increase in privacy of at least 70% is achievable with a decrease in accuracy of only 0.2% with the noise mechanism, while the hiding technique yields a decrease in accuracy of 11.2%.

In summary, we find that the noise mechanism presented in Section 6.3, providing epigeno-indistinguishability, is able to achieve a better privacy-utility trade-off than the hiding mechanism for the vast majority of studied diseases (17 out of 19). We have also shown in Section 6.2 that the privacy improvement with the hiding mechanism could actually be too optimistic due to the correlations existing between miRNAs. This is another argument to favor the noise mechanism rather than the hiding technique. Moreover, the p-values used to rank the miRNAs in the hiding mechanism actually require that, at some point in time, some entity, gets access to the full set of miRNAs of a significant number of individuals in order to measure these p-values. The noise mechanism is fully distributed and does not need to rely on a trusted entity at any point in time. Finally, it allows for more flexibility as it enables, e.g., the biomedical research community to access all miRNA expression levels of contributors.

7 Related Work

We start with the literature highlighting new privacy issues stemming from various types of biomedical data. Schadt et al. have shown that RNA expression data could be used to accurately predict genotypes [48]. The au-

thors present a Bayesian framework that relies on the association existing between expression levels of thousands of genes and genomic variations called expression quantitative trait loci (eQTLs). In the same vein, Philibert et al. demonstrate how methylation array data can be used to construct individually identifying genetic profiles, and to infer substance-use histories, such as alcohol or smoking [45]. Dyke et al. also study privacy risks related to methylation data, and discuss various methods to balance data open-access and (epi)genomic privacy [18]. Franzosa et al. evaluate how different samples of human microbiomes can be linked over time [23]. Their results show that more than 80% of individuals can still be uniquely identified one year later. Fierer et al. had already provided some evidence on the feasibility of linking skin bacterial communities back in 2010, but with very few individuals [22].

There has been quite a lot of work on determining membership of individuals in datasets, which is different from linking them over time among different datasets. Also, these previous works focus on genomic data only. Specifically, the attack aims to identify a victim's participation in a genome-wide association study (GWAS) based on aggregate statistics on the GWAS dataset, knowing the victim's genome (or part of it). Homer et al. are the first to thoroughly assess the feasibility and robustness of such an attack by relying upon statistics such as allele frequency or genotype counts [27]. Wang et al. extend the initial attack by making use of the correlations among the different positions in the genome [52]. Their attack proves to be effective with the statistics related to only a few hundreds genetic variants. Im et al. show that, if the victim's phenotype is rather extreme or if multiple phenotypes are available, regression coefficients can reveal the victim's participation in a genome-wide association study as much as allele frequencies [31]. Dwork et al. have very recently demonstrated the robustness of such an attack on distorted summary statistics [17].

On the protection side, various papers have studied how to apply noise to summary statistics to protect the privacy of GWAS participants. Johnson and Shmatikov design and implement algorithms for accurate and differentially private computation of various statistics of interest, such as the location of the most significant genomic variants, or the p-values of statistical tests between a given variant and the associated diseases [32]. Uhler et al. have also proposed to rely upon differential privacy for sharing GWAS results privately. In [51], they present methods for privately disclosing allele frequencies, chi-square statistics, and p-values. In [54], Yu et al. extend these methods by allowing for arbitrary number of cases and controls, assess their performance and compare it with the mechanism proposed by Johnson and Shmatikov. In [55], Yu et al. present a differentially-private mechanism for logistic regression and show how it can be applied to the analysis of GWAS data. In the pharmacogenetic context, Fredrikson et al. show that differential privacy mechanisms can induce bad warfarin dosing, thus expose patients to increased risk of stroke, bleeding events, and mortality [24]. Tramèr et al. [50] investigate how a relaxation of differential privacy that considers more reasonable amounts of background knowledge can help reach a better privacy-utility trade-off for releasing differentially private chi-square statistics in GWAS.

Our work differs from these in the sense that one of our protection mechanisms directly applies noise on the raw miRNA data to guarantee a certain degree of indistinguishability between them, instead of adding noise to summary statistics to ensure differential privacy. Our second defense technique relies on sharing a subset of miRNA data, which is closer to what Humbert et al. have developed in the genomic-privacy context. In particular, they propose an optimization algorithm that enables to share raw genomic variants (rather than summary statistics), e.g., for research, satisfying the genomic privacy requirements of all individuals in a family [29]. More generally, our work aims to protect real-valued miRNA expression vectors, which vary over time much more than DNA data.

8 Conclusion

To the best of our knowledge, this work is the very first to demonstrate that personal miRNA expression profiles can be successfully tracked over time. Our study sheds light on a widely overlooked problem, namely privacy risks stemming from epigenetic data, and brings this issue to the attention of both the biomedical and computer security research communities. In addition to the in-depth evaluation of the temporal linkability of miRNA expression profiles, we propose two defense mechanisms based on well-established privacy-enhancing methods: (i) hiding a subset of the expression data, and (ii) adding noise to the released expression profiles. We thoroughly evaluate the impact of these countermeasures on biomedical utility by studying how much accuracy decrease they induce in a typical machine-learning algorithm for predicting diseases. We observe that, for the majority of the 19 diseases studied in our experiments, the noise mechanism provides a better privacy-utility trade-off than the hiding method. Moreover, we highlight that the noise mechanism can be applied directly by the data contributors, independently of other contributors, and provides more flexibility for the biomedical community. Our work demonstrates that achieving indistinguishability by adding noise is a promising technique that could be applied to other types of biomedical data in the future.

Our results provide enough evidence about the extent of the threat to remove miRNA expression data from publicly accessible databases. Due to the limited number of individuals present in our datasets, we could not rely on supervised learning algorithms, which would certainly further improve the tracking capabilities of the adversary. We hope that this work will lead to further research on better understanding and protecting the privacy of miRNA expression data. Considering larger databases or uncertain membership of participants in the targeted databases are other promising directions for follow-up work.

9 Acknowledgements

This work has been partially funded by the German Research Foundation (DFG) via the collaborative research center "Methods and Tools for Understanding and Controlling Privacy" (SFB 1223), project A5.

References

[1] Arrayexpress. https://www.ebi.ac.uk/arrayexpress. Accessed: 2016-02-12.

[2] Bilans de santé en balade sur le net. http://www.lematin.ch/suisse/bilans-sante-balade-net/story/21621328. Accessed: 2016-02-03.

[3] The black market for stolen health care data. http://www.npr.org/sections/alltechconsidered/2015/02/13/385901377/the-black-market-for-stolen-health-care-data. Accessed: 2016-02-03.

[4] Gene expression omnibus. http://www.ncbi.nlm.nih.gov/geo. Accessed: 2016-02-12.

[5] Health insurer anthem discloses customer and employee data breach. http://www.computerworld.com/article/2879649/health-insurer-anthem-discloses-customer-and-employee-data-breach.html. Accessed: 2016-02-03.

[6] Medical data - a new target for hackers. https://www.logpoint.com/se/about-us/blog/249-medical-data-a-new-target-for-hackers. Accessed: 2016-02-03.

[7] micrornas: Definition and overview. https://www.thermofisher.com/de/de/home/references/ambion-tech-support/microrna-studies/tech-notes/micrornas-definition-and-overview.html. Accessed: 2016-02-12.

[8] Premera, anthem data breaches linked by similar hacking tactics. http://www.computerworld.com/article/2898419/data-breach/premera-anthem-data-breaches-linked-by-similar-hacking-tactics.html. Accessed: 2016-02-03.

[9] Urgent probe as michael schumacher's medical records stolen and put on sale for 40k. http://www.express.co.uk/news/world/484495/Investigation-underway-after-Michael-Schumacher-s-medical-records-stolen. Accessed: 2016-02-03.

[10] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 901–914. ACM, 2013.

[11] E. Ayday, E. De Cristofaro, J.-P. Hubaux, and G. Tsudik. Whole genome sequencing: Revolutionary medicine or privacy nightmare? *Computer*, pages 58–66, 2015.

[12] C. Backes, P. Leidinger, A. Keller, M. Hart, T. Meyer, E. Meese, and A. Hecksteden. Blood born mirnas signatures that can serve as disease specific biomarkers are not significantly affected by overall fitness and exercise. *PloS one*, 9(7):e102183, 2014.

[13] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

[14] C. Cachin. *Entropy measures and unconditional security in cryptography*. PhD thesis, SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH, 1997.

[15] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi. Broadening the scope of differential privacy using metrics. In *Privacy Enhancing Technologies*, pages 82–102. Springer, 2013.

[16] J. Cloud. Why your DNA isn't your destiny. *Time*, January 2010.

[17] C. Dwork, A. Smith, T. Steinke, J. Ullman, and S. Vadhan. Robust traceability from trace amounts. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 650–669. IEEE, 2015.

[18] S. O. Dyke, W. A. Cheung, Y. Joly, O. Ammerpohl, P. Lutsik, M. A. Rothstein, M. Caron, S. Busche, G. Bourque, L. Rönnblom, et al. Epigenome data release: a participant-centered approach to privacy protection. *Genome biology*, 16:1–12, 2015.

[19] J. Edmonds. Paths, trees, and flowers. *Canadian Journal of mathematics*, 17(3):449–467, 1965.

[20] Y. Erlich and A. Narayanan. Routes for breaching and protecting genetic privacy. *Nature Reviews Genetics*, 15:409–421, 2014.

[21] A. P. Feinberg and M. D. Fallin. Epigenetics at the crossroads of genes and the environment. *JAMA*, 314:1129–1130, 2015.

[22] N. Fierer, C. L. Lauber, N. Zhou, D. McDonald, E. K. Costello, and R. Knight. Forensic identification using skin bacterial communities. *Proceedings of the National Academy of Sciences*, 107(14):6477–6481, 2010.

[23] E. A. Franzosa, K. Huang, J. F. Meadow, D. Gevers, K. P. Lemon, B. J. Bohannan, and C. Huttenhower. Identifying personal microbiomes using metagenomic codes. *Proceedings of the National Academy of Sciences*, page 201423854, 2015.

[24] M. Fredrikson, E. Lantz, S. Jha, S. Lin, D. Page, and T. Ristenpart. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 17–32, 2014.

[25] R. C. Friedman, K. K.-H. Farh, C. B. Burge, and D. P. Bartel. Most mammalian mrnas are conserved targets of micrornas. *Genome research*, 19(1):92–105, 2009.

[26] M. Gymrek, A. L. McGuire, D. Golan, E. Halperin, and Y. Erlich. Identifying personal genomes by surname inference. *Science*, 339:321–324, 2013.

[27] N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, S. F. Nelson, and D. W. Craig. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS Genet*, 4(8):e1000167, 2008.

[28] M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti. Addressing the concerns of the Lacks family: quantification of kin genomic privacy. In *Proceedings of the 2013 ACM SIGSAC CCS*, pages 1141–1152, 2013.

[29] M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti. Reconciling utility with privacy in genomics. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 11–20. ACM, 2014.

[30] M. Humbert, K. Huguenin, J. Hugonot, E. Ayday, and J.-P. Hubaux. De-anonymizing genomic databases using phenotypic traits. *Proceedings on Privacy Enhancing Technologies(PoPETs)*, 2015.

[31] H. K. Im, E. R. Gamazon, D. L. Nicolae, and N. J. Cox. On sharing quantitative trait gwas results in an era of multiple-omics data and the limits of genomic privacy. *The American Journal of Human Genetics*, 90(4):591–598, 2012.

[32] A. Johnson and V. Shmatikov. Privacy-preserving data exploration in genome-wide association studies. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1079–1087. ACM, 2013.

[33] P. A. Jones and S. B. Baylin. The epigenomics of cancer. *Cell*, 128:683–692, 2007.

[34] A. Keller, P. Leidinger, A. Bauer, A. ElSharawy, J. Haas, C. Backes, A. Wendschlag, N. Giese, C. Tjaden, K. Ott, et al. Toward the blood-borne mirnome of human diseases. *Nature methods*, 8:841–843, 2011.

[35] A. Keller, P. Leidinger, B. Vogel, C. Backes, A. ElSharawy, V. Galata, S. C. Mueller, S. Marquart, M. G. Schrauder, R. Strick, et al. mirnas can be generally associated with human pathologies as exemplified for mir-144*. *BMC medicine*, 12(1):224, 2014.

[36] F. Koufogiannis, S. Han, and G. J. Pappas. Optimality of the laplace mechanism in differential privacy. *arXiv preprint arXiv:1504.00065*, 2015.

[37] P. Leidinger, C. Backes, S. Deutscher, K. Schmitt, S. C. Mueller, K. Frese, J. Haas, K. Ruprecht, F. Paul, C. Stahler, et al. A blood based 12-mirna signature of alzheimer disease patients. *Genome Biol*, 14:R78, 2013.

[38] P. Leidinger, V. Galata, C. Backes, C. Stähler, S. Rheinheimer, H. Huwer, E. Meese, and A. Keller. Longitudinal study on circulating mirnas in patients after lung cancer resection. *Oncotarget*, 6:16674, 2015.

[39] Z. Lin, A. B. Owen, and R. B. Altman. Genomic research and human subject privacy. *SCIENCE-NEW YORK THEN WASHINGTON-.*, pages 183–183, 2004.

[40] J. Lu, G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, et al. Microrna expression profiles classify human cancers. *nature*, 435(7043):834–838, 2005.

[41] J. L. Massey. Guessing and entropy. In *Information Theory, 1994. Proceedings., 1994 IEEE International Symposium on*, page 204. IEEE, 1994.

[42] M. Naveed, E. Ayday, E. W. Clayton, J. Fellay, C. A. Gunter, J.-P. Hubaux, B. A. Malin, and X. Wang. Privacy in the genomic era. *ACM Computing Surveys (CSUR)*, 48:6, 2015.

[43] T. Ngun et al. Abstract: A novel predictive model of sexual orientation using epigenetic markers. In *American Society of Human Genetics 2015 Annual Meeting*, 2015.

[44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[45] R. A. Philibert, N. Terry, C. Erwin, W. J. Philibert, S. R. Beach, and G. H. Brody. Methylation array data can simultaneously identify individuals and convey protected health information: an unrecognized ethical concern. *Clinical epigenetics*, 6:28, 2014.

[46] I. A. Qureshi and M. F. Mehler. Advances in epigenetics and epigenomics for neurodegenerative diseases. *Current neurology and neuroscience reports*, 11:464–473, 2011.

[47] M. A. Rothstein, Y. Cai, and G. E. Marchant. The ghost in our genes: legal and ethical implications of epigenetics. *Health matrix (Cleveland, Ohio: 1991)*, 19:1, 2009.

[48] E. E. Schadt, S. Woo, and K. Hao. Bayesian method to predict individual snp genotypes from gene expression data. *Nature genetics*, 44:603–608, 2012.

[49] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

[50] F. Tramèr, Z. Huang, J.-P. Hubaux, and E. Ayday. Differential privacy with bounded priors: reconciling utility and privacy in genome-wide association studies. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 1286–1297. ACM, 2015.

[51] C. Uhler, A. Slavković, and S. E. Fienberg. Privacy-preserving data sharing for genome-wide association studies. *The Journal of privacy and confidentiality*, 5(1):137, 2013.

[52] R. Wang, Y. F. Li, X. Wang, H. Tang, and X. Zhou. Learning your identity and disease from research papers: information leaks in genome wide association study. In *Proceedings of the 16th ACM conference on Computer and communications security*, pages 534–544. ACM, 2009.

[53] L. D. Wood, D. W. Parsons, S. Jones, J. Lin, T. Sjöblom, R. J. Leary, D. Shen, S. M. Boca, T. Barber, J. Ptak, et al. The genomic landscapes of human breast and colorectal cancers. *Science*, 318:1108–1113, 2007.

[54] F. Yu, S. E. Fienberg, A. B. Slavković, and C. Uhler. Scalable privacy-preserving data sharing methodology for genome-wide association studies. *Journal of biomedical informatics*, 50:133–141, 2014.

[55] F. Yu, M. Rybar, C. Uhler, and S. E. Fienberg. Differentially-private logistic regression for detecting multiple-snp association in gwas databases. In *Privacy in Statistical Databases*, pages 170–184. Springer, 2014.

A Appendix

A.1 Human Subjects and Ethical Considerations

The studies have received an approval from our institutional ethics review board. Moreover, not only have all datasets been stored and analyzed in anonymized form, but we also handled our results with great care to not deanonymize any of the patients. This way, we ensured that all participants were treated equally and with respect.

$\Delta_{\text{acc}}^{\text{max}}$	0.5%		1.0%		2.0%		3.0%		4.0%		5.0%	
Disease	Δ_{priv}^{m}	$\Delta_{\text{priv}}^{\varepsilon}$	Δ_{priv}^{m}	$\Delta_{\text{priv}}^{\varepsilon}$	Δ_{priv}^{m}	$\Delta_{\text{priv}}^{\varepsilon}$	Δ_{priv}^{m}	$\Delta_{\text{priv}}^{\varepsilon}$	Δ_{priv}^{m}	$\Delta_{\text{priv}}^{\varepsilon}$	Δ_{priv}^{m}	$\Delta_{\text{priv}}^{\varepsilon}$
Periodontitis	26.9%	74.1%	26.9%	79.2%	50.0%	79.2%	88.5%	83.6%	88.5%	83.6%	88.5%	83.6%
Renal cancer	30.8%	-	30.8%	3.6%	69.2%	5.2%	73.1%	60.8%	73.1%	72.7%	80.8%	78.8%
Wilms tumor	3.8%	6.4%	7.7%	9.5%	7.7%	40.1%	7.7%	61.5%	7.7%	70.4%	11.5%	74.3%
Benign prostate hyperplasia	−3.8%	10.6%	3.8%	70.5%	11.5%	72.2%	46.2%	79.2%	57.7%	79.2%	65.4%	79.2%
Chronic obstructive pulmonary disease (COPD)	0.0%	2.7%	0.0%	5.5%	0.0%	12.5%	0.0%	12.5%	15.4%	50.3%	23.1%	69.8%
Colon cancer	19.2%	11.4%	30.8%	30.5%	30.8%	60.2%	57.7%	60.2%	73.1%	70.5%	73.1%	73.8%
Ductal adenocarcinoma	0.0%	50.6%	3.8%	50.6%	7.7%	62.5%	42.3%	62.5%	50.0%	69.5%	50.0%	74.2%
Glioma	65.4%	5.2%	65.4%	5.2%	80.8%	68.5%	80.8%	80.8%	80.8%	87.5%	80.8%	87.5%
Lung cancer	11.5%	74.1%	46.2%	79.3%	50.0%	79.3%	50.0%	79.3%	50.0%	79.3%	50.0%	79.3%
Melanoma	0.0%	-	0.0%	3.8%	0.0%	5.9%	3.8%	10.3%	38.5%	40.2%	38.5%	60.7%
Multiple sclerosis	19.2%	49.5%	53.8%	62.6%	53.8%	62.6%	61.5%	62.6%	61.5%	73.7%	61.5%	73.7%
Myocardial infarction	3.8%	52.4%	3.8%	52.4%	3.8%	60.5%	38.5%	60.5%	42.3%	74.6%	42.3%	74.6%
Non-ischaemic systolic heart failure	−3.8%	80.0%	0.0%	80.0%	46.2%	80.0%	46.2%	84.7%	46.2%	84.7%	46.2%	84.7%
Ovarian cancer	26.9%	78.5%	26.9%	78.5%	42.3%	78.5%	42.3%	84.3%	50.0%	84.3%	50.0%	86.2%
Pancreatitis	19.2%	10.3%	26.9%	39.8%	26.9%	53.5%	26.9%	53.5%	57.7%	62.2%	65.4%	71.2%
Prostate cancer	−3.8%	-	−3.8%	-	3.8%	4.0%	42.3%	6.2%	42.3%	10.1%	42.3%	38.5%
Psoriasis	0.0%	6.5%	0.0%	31.4%	3.8%	74.0%	19.2%	80.1%	23.1%	80.1%	61.5%	80.1%
Sarcoidosis	0.0%	69.3%	3.8%	74.0%	50.0%	79.8%	92.3%	79.8%	92.3%	79.8%	92.3%	79.8%
Tumor of stomach	15.4%	69.8%	34.6%	69.8%	65.4%	79.4%	65.4%	79.4%	65.4%	84.6%	65.4%	84.6%

Table 2: Relative increase in privacy for both defense mechanisms in relation to a fixed maximal decrease in accuracy. "-" means that the respective maximal decrease in accuracy was not achievable with any ε we tested for. A negative value means that the attack's success rate could, in this case, even exceed the success rate with all miRNAs taken into account.

$\Delta^{\text{min}}_{\text{priv}}$	30.0%		40.0%		50.0%		60.0%		70.0%		80.0%	
Disease	Δ^{m}_{acc}	$\Delta^{\varepsilon}_{\text{acc}}$	Δ^{m}_{acc}	$\Delta^{\varepsilon}_{\text{acc}}$	Δ^{m}_{acc}	$\Delta^{\varepsilon}_{\text{acc}}$	Δ^{m}_{acc}	$\Delta^{\varepsilon}_{\text{acc}}$	Δ^{m}_{acc}	$\Delta^{\varepsilon}_{\text{acc}}$	Δ^{m}_{acc}	$\Delta^{\varepsilon}_{\text{acc}}$
Periodontitis	1.9%	−1.9%	1.9%	−1.9%	2.6%	−1.9%	2.6%	−1.9%	2.6%	−0.8%	2.6%	2.9%
Renal cancer	0.0%	2.3%	1.7%	2.3%	1.7%	2.3%	1.7%	2.3%	2.5%	3.1%	4.8%	7.0%
Wilms tumor	5.2%	1.4%	5.2%	1.7%	5.5%	2.2%	5.5%	2.8%	8.1%	3.2%	15.5%	11.3%
Benign prostate hyperplasia	2.7%	0.5%	2.7%	0.6%	3.5%	0.6%	5.0%	0.9%	5.6%	0.9%	5.6%	5.5%
Chronic obstructive pulmonary disease (COPD)	7.9%	3.3%	12.0%	3.3%	12.0%	3.3%	15.4%	4.1%	15.6%	5.3%	15.6%	9.0%
Colon cancer	0.7%	0.8%	2.4%	1.3%	2.4%	1.3%	3.3%	1.9%	3.9%	3.3%	7.7%	9.4%
Ductal adenocarcinoma	2.8%	0.1%	2.8%	0.4%	5.2%	0.4%	5.2%	1.8%	6.4%	4.6%	6.4%	6.4%
Glioma	0.0%	1.2%	0.0%	1.2%	0.4%	1.2%	0.4%	1.4%	1.1%	2.1%	1.1%	2.8%
Lung cancer	0.7%	−1.5%	1.0%	−1.5%	6.6%	−1.5%	8.1%	−1.5%	11.2%	0.2%	18.2%	5.5%
Melanoma	3.7%	3.4%	5.0%	3.4%	7.5%	4.1%	7.5%	4.9%	7.5%	6.5%	10.0%	11.2%
Multiple sclerosis	0.9%	−0.0%	0.9%	0.1%	0.9%	0.7%	2.3%	0.7%	8.1%	3.8%	8.1%	6.7%
Myocardial infarction	2.8%	0.0%	3.6%	0.4%	7.2%	0.4%	7.3%	1.3%	7.3%	3.3%	11.2%	6.7%
Non-ischaemic systolic heart failure	2.0%	−2.6%	2.0%	−2.6%	8.5%	−2.1%	8.5%	−2.1%	9.3%	−1.5%	9.3%	2.5%
Ovarian cancer	1.3%	−0.7%	1.3%	−0.7%	5.5%	−0.7%	6.7%	−0.7%	9.0%	−0.7%	9.0%	2.5%
Pancreatitis	3.8%	0.8%	3.8%	1.9%	3.8%	1.9%	4.5%	3.1%	7.9%	4.3%	7.9%	7.9%
Prostate cancer	2.7%	4.8%	2.7%	5.0%	7.6%	5.0%	7.6%	5.6%	7.6%	5.6%	11.5%	8.9%
Psoriasis	4.3%	1.0%	4.3%	1.3%	4.3%	1.3%	4.3%	1.4%	5.8%	1.4%	10.0%	2.1%
Sarcoidosis	1.4%	−0.2%	1.6%	−0.2%	2.2%	−0.2%	2.2%	−0.2%	2.2%	0.6%	2.2%	5.3%
Tumor of stomach	0.9%	−0.0%	1.7%	−0.0%	2.0%	−0.0%	2.0%	−0.0%	5.1%	1.1%	5.1%	3.3%

Table 3: Relative decrease in accuracy for both defense mechanisms in relation to a fixed minimal increase in privacy. A negative value means that the accuracy could, in this case, even exceed the baseline accuracy (utility).