How to Keep your Genome Secret

Homomorphic Encryption for Private Genomic Predictions

KRISTIN LAUTER
CRYPTOGRAPHY RESEARCH GROUP
MICROSOFT RESEARCH

AMS-MAA INVITED TALK

JANUARY 8, 2016

Protecting Data via Encryption:
Homomorphic encryption

1. Put your gold in a locked box.
2. Keep the key.
3. Let your jeweler work on it through a glove box.
4. Unlock the box when the jeweler is done!

Homomorphic Encryption: addition

$$\begin{tikzcd}
a, b \arrow[r, "\text{compute}"'] \arrow[d, "\text{encrypt}"'] & a+ b \arrow[d, "\text{encrypt}"] \\
E(a), E(b) \arrow[r, "\text{compute}"'] & {}
\end{tikzcd}$$

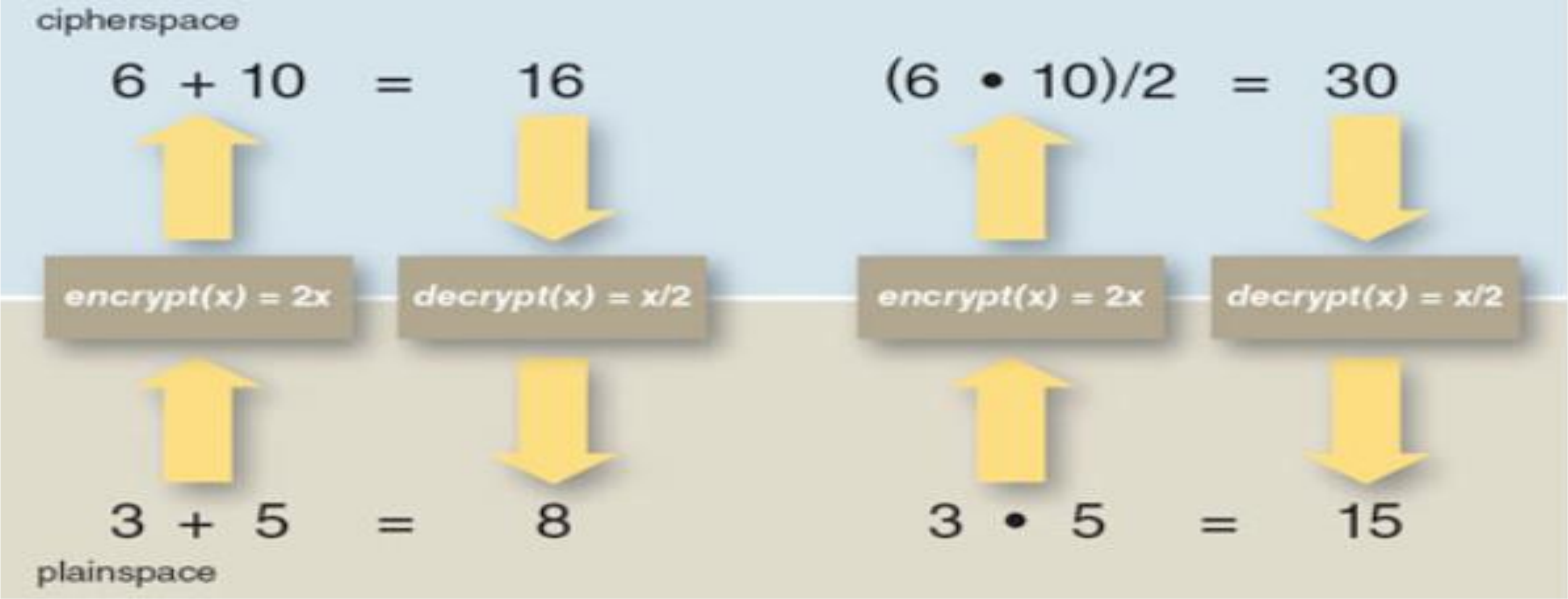
E(a) + E(b) = E(a+b)

Homomorphic Encryption: multiplication

$$\begin{tikzcd}
a, b \arrow[r, "\text{compute}"'] \arrow[d, "\text{encrypt}"'] & a \times b \arrow[d, "\text{encrypt}"] \\
E(a), E(b) \arrow[r, "\text{compute}"'] & {}
\end{tikzcd}$$

$E(a) \times E(b) = E(ab)$

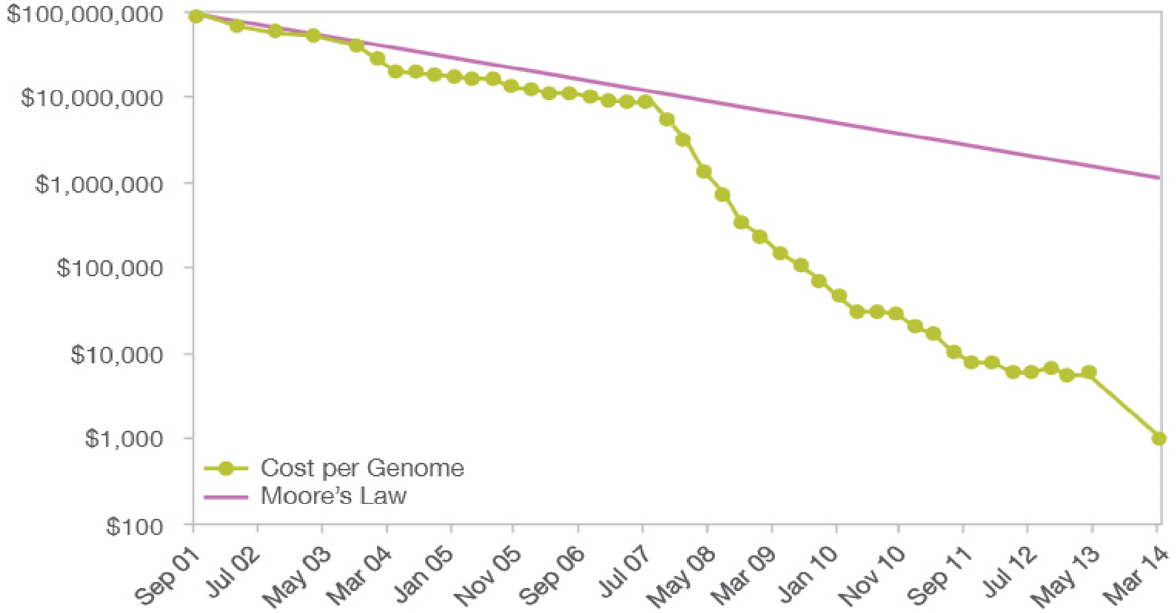
Homomorphic encryption



American Scientist, Sept/Oct 2012

Genomic Revolution

- Fast drop in the cost of genome-sequencing
  - 2000: $3 billion
  - Mar. 2014: $1,000
  - Genotyping 1M variations: below $200



- Unleashing the potential of the technology
  - Healthcare: e.g., disease risk detection, personalized medicine
  - Biomedical research: e.g., geno-phono association
  - Legal and forensic
  - DTC: e.g., ancestry test, paternity test
  - Million Veterans Program

……

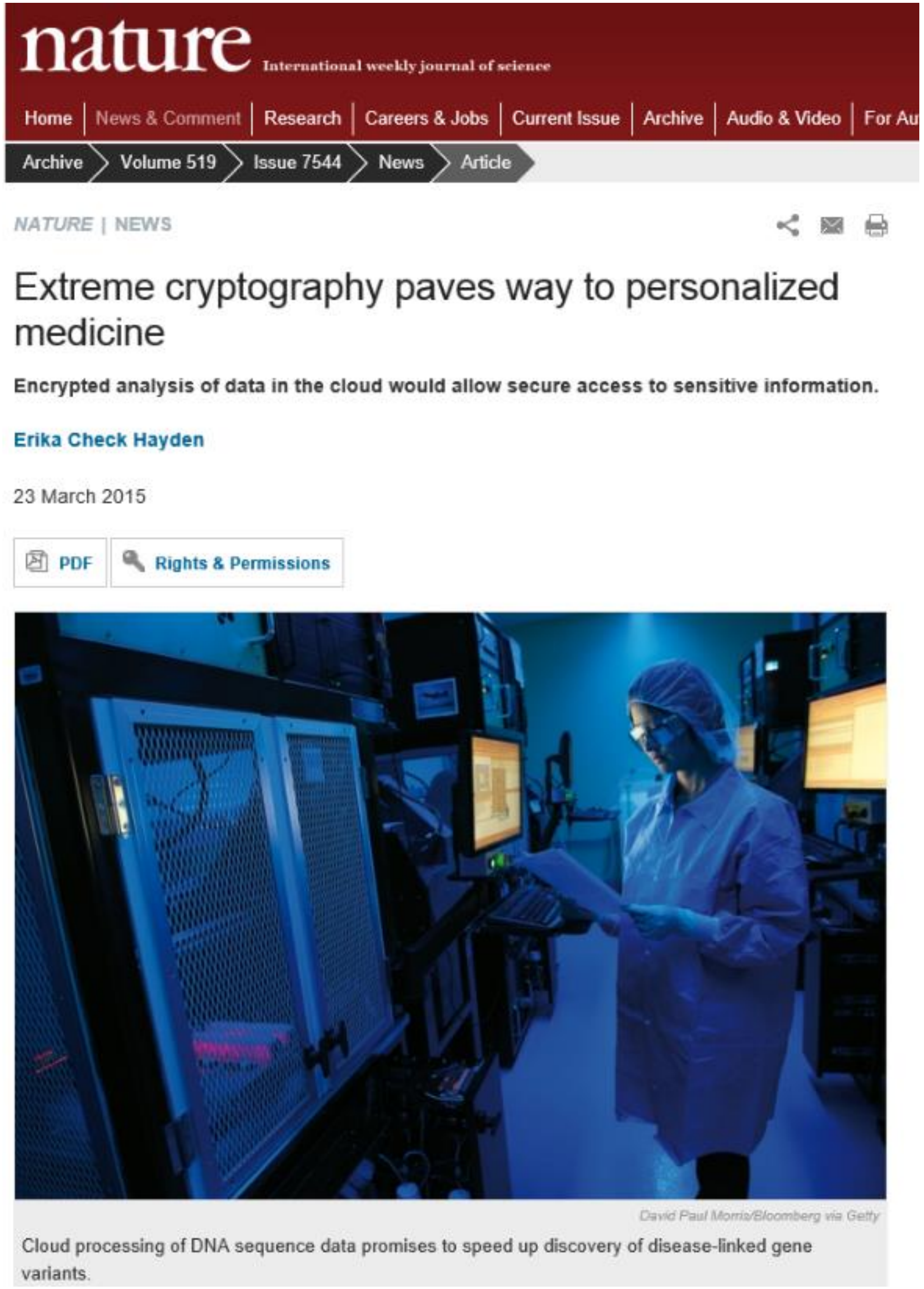
Genome Privacy

- Privacy risks
  - Genetic disease disclosure
  - Collateral damage
  - Genetic discrimination

- Grand Challenges:
  - **How to share genomic data or learning in a way that preserves the privacy of the data donors, without undermining the utility of the data or impeding its convenient dissemination?**
  - **How to perform LARGE-SCALE, PRIVACY-PRESERVING analysis on genomic data, in an untrusted cloud environment or across multiple users?**

Secure Genome Analysis Competition

- iDASH Privacy & Security Workshop 2015
- Sponsored by NIH (National Institutes of Health)
- Submission deadline: Feb 28 2015
- Workshop: March 16, 2015
  UCSD Medical Education and Telemedicine Building
- Media coverage in GenomeWeb, Donga Science, Nature
- Teams from: Microsoft, IBM, Stanford/MIT, UCI, University of Tsukuba, …
- Two Tracks: Multi-Party Computation and Homomorphic Encryption
- Challenges: GWAS and Sequence Alignment


nature
International weekly journal of science
Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Au
Archive > Volume 519 > Issue 7544 > News > Article
PDF
Rights & Permissions

NATURE | NEWS

Extreme cryptography paves way to personalized medicine

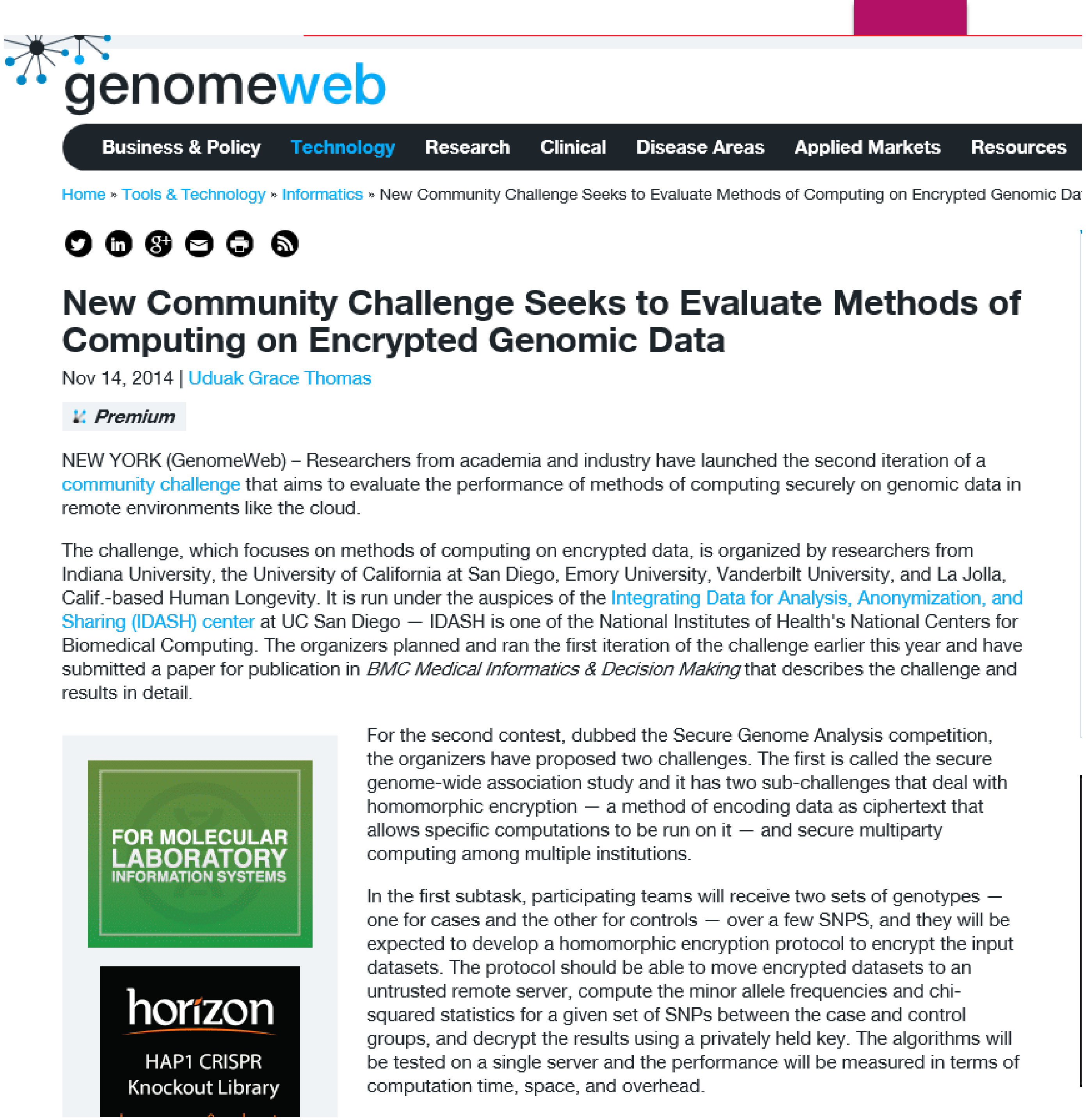
Encrypted analysis of data in the cloud would allow secure access to sensitive information.

Erika Check Hayden

23 March 2015

David Paul Morris/Bloomberg via Getty

Cloud processing of DNA sequence data promises to speed up discovery of disease-linked gene variants.


genomeweb
Business & Policy | Technology | Research | Clinical | Disease Areas | Applied Markets | Resources
Home » Tools & Technology » Informatics » New Community Challenge Seeks to Evaluate Methods of Computing on Encrypted Genomic Da
Premium

New Community Challenge Seeks to Evaluate Methods of Computing on Encrypted Genomic Data

Nov 14, 2014 | Uduak Grace Thomas

NEW YORK (GenomeWeb) – Researchers from academia and industry have launched the second iteration of a community challenge that aims to evaluate the performance of methods of computing securely on genomic data in remote environments like the cloud.

The challenge, which focuses on methods of computing on encrypted data, is organized by researchers from Indiana University, the University of California at San Diego, Emory University, Vanderbilt University, and La Jolla, Calif.-based Human Longevity. It is run under the auspices of the Integrating Data for Analysis, Anonymization, and Sharing (IDASH) center at UC San Diego — IDASH is one of the National Institutes of Health's National Centers for Biomedical Computing. The organizers planned and ran the first iteration of the challenge earlier this year and have submitted a paper for publication in *BMC Medical Informatics & Decision Making* that describes the challenge and results in detail.

For the second contest, dubbed the Secure Genome Analysis competition, the organizers have proposed two challenges. The first is called the secure genome-wide association study and it has two sub-challenges that deal with homomorphic encryption — a method of encoding data as ciphertext that allows specific computations to be run on it — and secure multiparty computing among multiple institutions.

In the first subtask, participating teams will receive two sets of genotypes — one for cases and the other for controls — over a few SNPS, and they will be expected to develop a homomorphic encryption protocol to encrypt the input datasets. The protocol should be able to move encrypted datasets to an untrusted remote server, compute the minor allele frequencies and chi-squared statistics for a given set of SNPs between the case and control groups, and decrypt the results using a privately held key. The algorithms will be tested on a single server and the performance will be measured in terms of computation time, space, and overhead.

Donga Science, March 13, 2015

○ **MS 연구진 이끌고 DNA 보안 알고리즘 개발**

이 연구원과 같은 연구실에서 한솥밥을 먹고 있는 김미란 연구원(28)은 생체정보 보안 연구 분야에서 떠오르는 샛별이다. 그는 1월 미국 마이크로소프트(MS) 연구소 초청으로 현지에 급파됐다. 작년 내내 MS 연구진을 이끌고 개발한 DNA 보안 기술이 '안전 게놈 분석 경진대회(Secure Genome Analysis Competition)'에 출전했기 때문이다. 이 대회는 샌디에이고 캘리포니아대 의대가 지난해부터 개최하는 첨단 생체정보 보안 대회다.
http://news.donga.com/It/3/all/20150313/70100744/1

GenomeWeb, Nature, …

청불인 듯 청불 아닌 청불 같은 너에게서 흥행의 향기가…

외화 '킹스맨' 인기로 본 '청불 영화의 흥행 법칙'

[illegible]

[illegible]

한국영화는 범죄물이 단연 우세 역대 10위권에 6편 올라

[illegible]

서울아트시네마 '낙원에 고하는 작별인사'

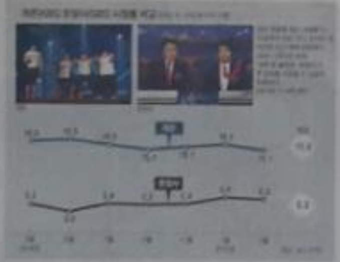
[illegible]

[illegible]

[illegible]

일요일 밤으로 옮기는 웃찾사, 개콘에 "한판 붙자"

[illegible]



[illegible]

최신 함수암호 무력화, 학계에 신선한 충격

[illegible]

[illegible]

최신 암호체계 허무는 방법 발표

최대 학술대회 '유로크립트'서

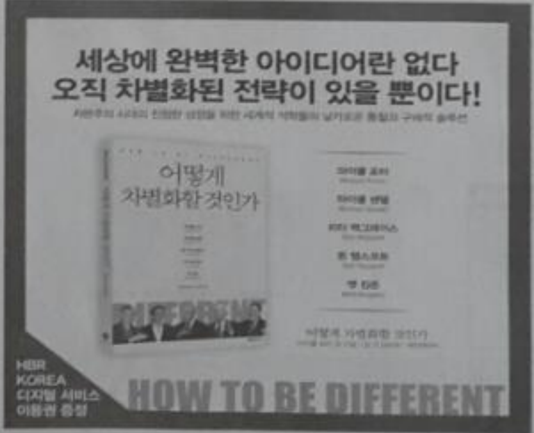
최우수 논문상 받아 이목 집중

DNA 보안 알고리즘 발표

[illegible]

베일벗은 달 지각활동 예상보다 훨씬 역동적

[illegible]



Why the excitement?

Fundamental Problem: privacy protection

- Burgeoning genome sequencing capability
- Explosion of scientific research possible
- High risk for personal privacy

Fundamental Progress through interaction

- Computer Scientists
- Mathematicians
- Bioinformaticians
- Policy-makers

Two Challenges!

Challenge 1:

Homomorphic encryption (HE) based secure genomic data analysis

- Task 1: Secure Outsourcing GWAS
- Task 2: Secure comparison between genomic data

Challenge 2:

Secure multiparty computing (MPC) based secure genomic data analysis

(two institutions)

- Task 1: Secure distributed GWAS
- Task 2: Secure comparison between genomic data

Data Source

- 200 Cases from Personal Genome Project (PGP)

PGP: http://www.personalgenomes.org/ launched by Harvard Medical School

- 200 Controls were simulated based on the haplotypes of 174 individuals from population of International HapMap Project (http://hapmap.ncbi.nlm.nih.gov/)

- 2 individual genomes (hu604D39 with 4,542,542 variations and hu661AD0 with 4,368,847 variations comparing to the reference human genome) were randomly selected from PGP

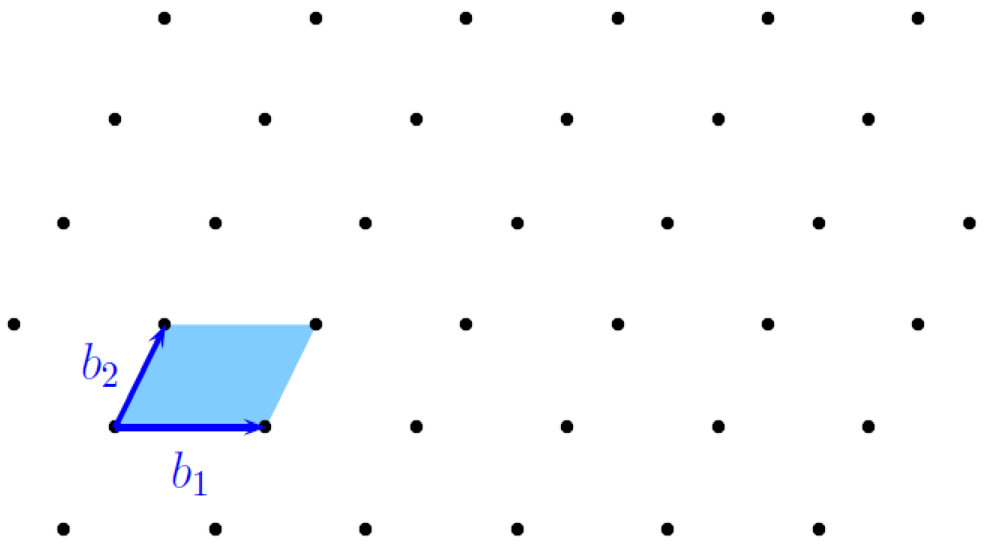
Contest Outcomes

- 11 teams, many winners!
- Microsoft Research won Edit Distance Task for Homomorphic Encryption
- Report to NIH
- Showed practical nature of computing on encrypted genomic data
- Influenced NIH policy recommendations for handling data in the cloud
- Special Issue in Biomedical Informatics and Medical Decision-making
  - Papers from each team describing their submissions
- NSF could sponsor such contests/publications on key challenges!

Mathematics of Homomorphic Encryption

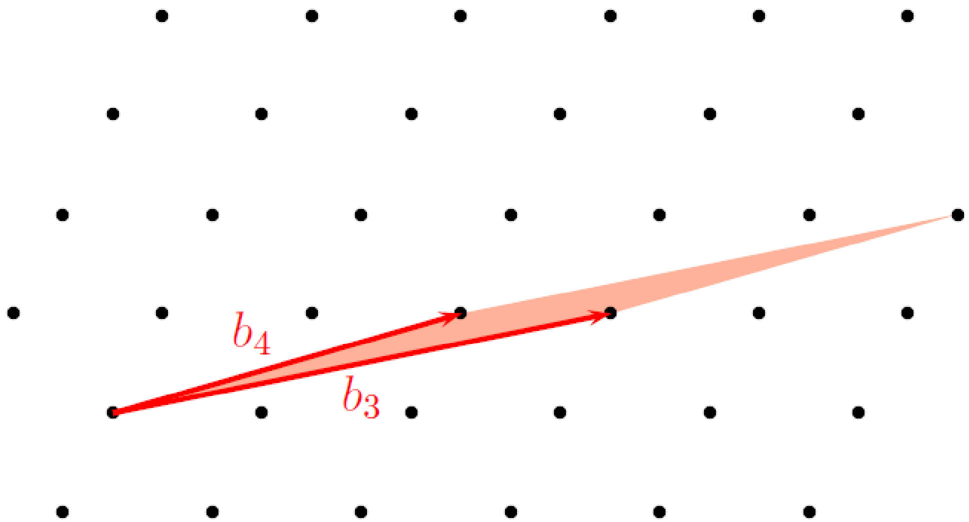
- New hard problems proposed (2009-2013), related to well-known hard lattice problems
  - Small Principal Ideal Problem, Approximate GCD, Learning With Errors (LWE), Ring-Learning With Errors
- Lattice-based Cryptography:
  - Compare to other public key systems: RSA (1975), ECC (1985), Pairings (2000)
  - Proposed by Hoffstein, Pipher, and Silverman in 1996 (NTRU), Aijtai-Dwork
- Hard Lattice Problems:
  - approximate Shortest Vector Problem, Bounded Distance Decoding
- SECURITY:
  - best attacks take exponential time
  - secure against quantum attacks (so far…)

Lattice with a Good (short) Basis



$$L = \mathbb{Z}b_1 + \mathbb{Z}b_2$$

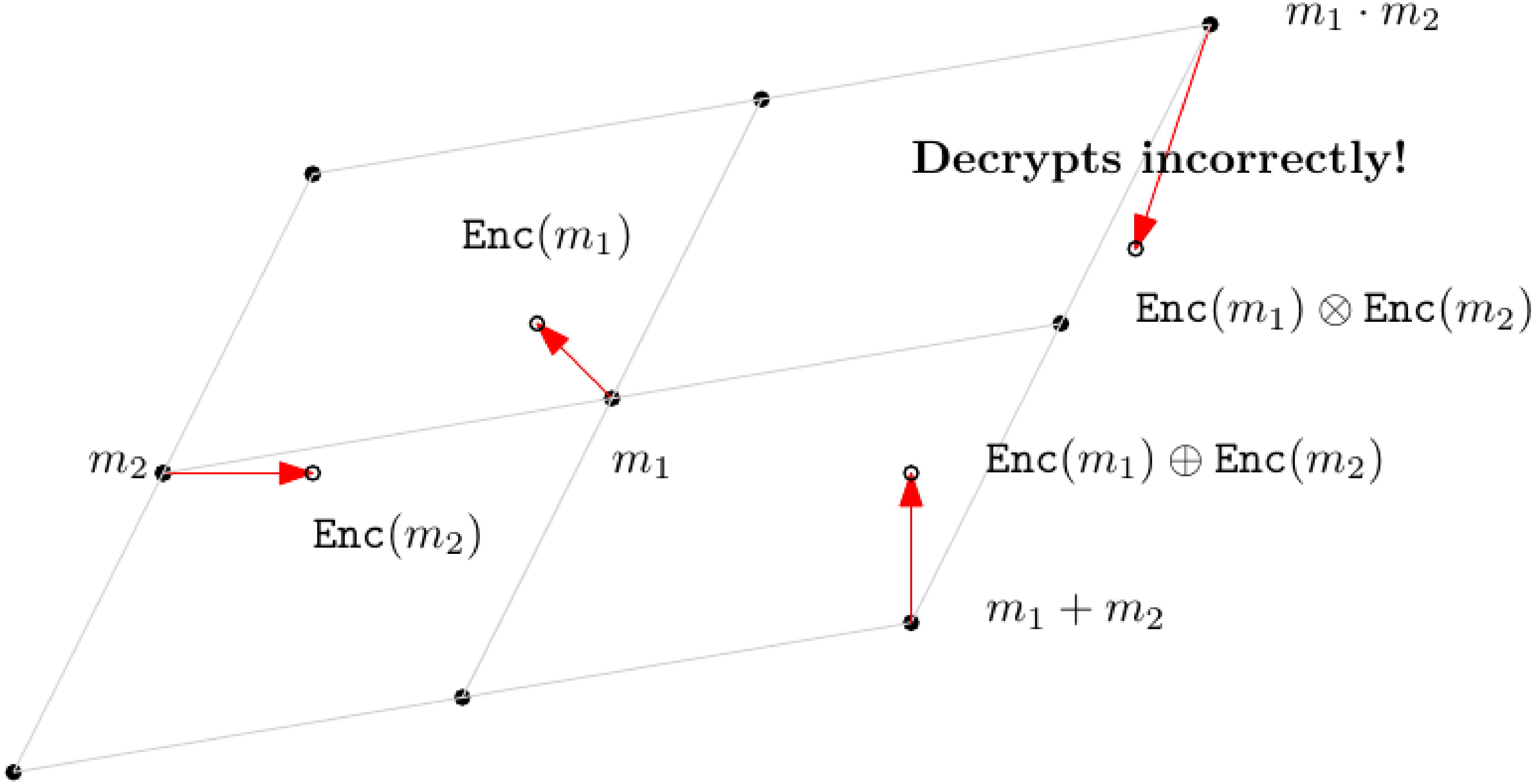
Lattice with a Bad Basis



$$L = \mathbb{Z}b_3 + \mathbb{Z}b_4$$

Idea of new schemes

- Lattice vectors → coefficients of polynomials
- Polynomials can be added and multiplied
- Encryption adds noise to a “secret” inner product
- Decryption subtracts the secret and then the noise becomes easy to cancel
- Hard problem is to “decode” noisy vectors
- If you have a short basis, it is easy to decompose vectors


$m_1 \cdot m_2$
Decrypts incorrectly!
$\texttt{Enc}(m_1)$
$\texttt{Enc}(m_1) \otimes \texttt{Enc}(m_2)$
m_2
m_1
$\texttt{Enc}(m_1) \oplus \texttt{Enc}(m_2)$
$\texttt{Enc}(m_2)$
$m_1 + m_2$

Decrypt by recovering the nearest lattice point using secret key information.

Ring-Learning With Errors (R-LWE)

- Let $q \equiv 1 \bmod 2n$ be a prime, $\mathbb{Z}_q = \mathbb{Z}/q\mathbb{Z}$. n=$2^k$. Consider the polynomial ring

$$R_q = \mathbb{Z}_q[x]/(x^n + 1).$$

- Given a secret element $s \in R_q$ and a number of pairs

$$(a_i, b_i = a_i s + e_i),$$

- where $a_i \leftarrow R_q$ are chosen uniformly at random, and $e_i \leftarrow D_\sigma(R_q)$ are chosen coefficientwise according to the discrete Gaussian error distribution $D_\sigma(\mathbb{Z}_q)$.

- **R-LWE problem:** Find the secret s (search), or distinguish whether a list of pairs (a_i, b_i) was chosen as described above or whether both $a_i, b_i \leftarrow R_q$were chosen uniformly at random (decision).

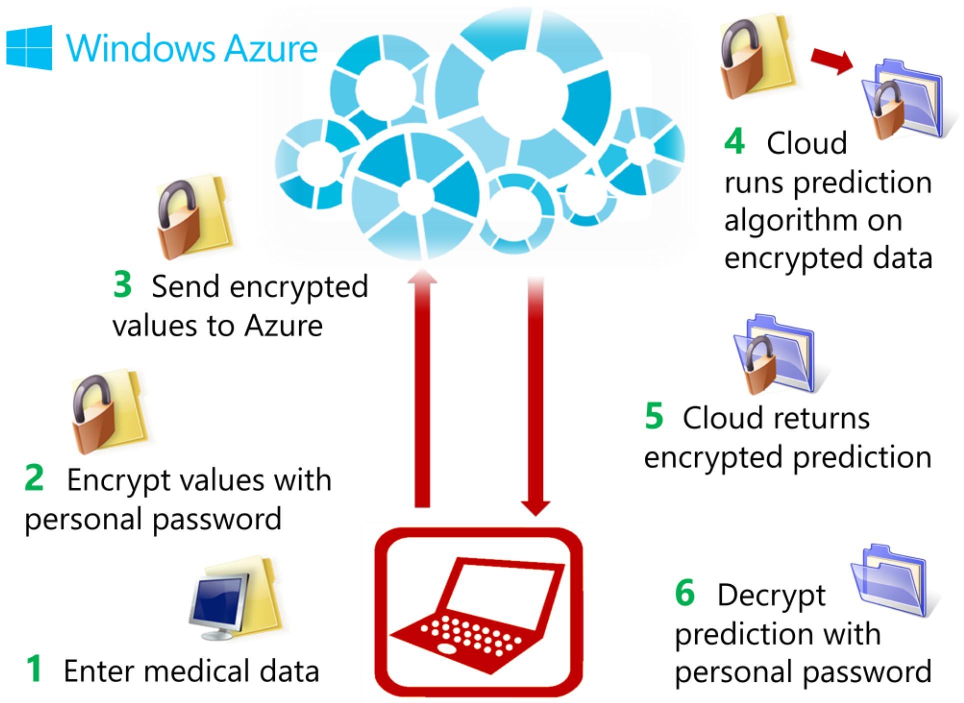
New questions in number theory

- Are these problems hard for other number rings??
- In general, NO
  - Eisentraeger-Hallgren-L (2014) + Elias-L-Ozman-Stange (2015)
- Questions:
  - distributions of elements of small order in finite fields,
  - relationship with Mahler measure,
  - construction of number rings with certain properties.

Scenarios: Private cloud services

- Direct-to-patient services
  - Personalized medicine
  - DNA sequence analysis
  - Disease prediction
- Hosted databases for enterprise
  - Hospitals, clinics, companies
  - Allows for third party interaction

Outsourcing computation



Windows Azure

1 Enter medical data

2 Encrypt values with personal password

3 Send encrypted values to Azure

4 Cloud runs prediction algorithm on encrypted data

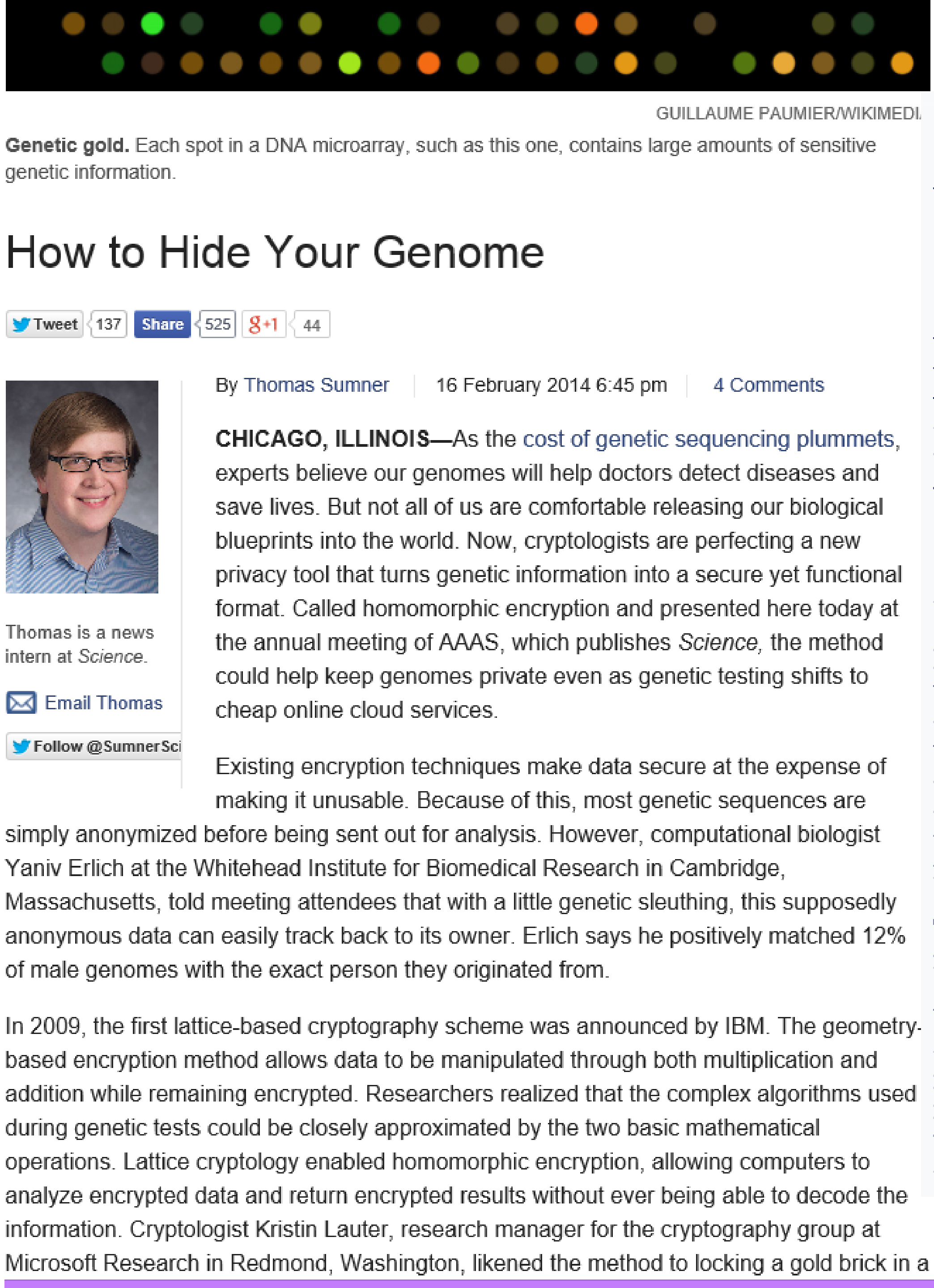
5 Cloud returns encrypted prediction

6 Decrypt prediction with personal password

Demo: Will you have a heart attack?

- Online service running in Windows Azure
- Patient enters personal info on local machine:
  - weight, age, height, blood pressure, body mass index
- Data is encrypted on local machine
- Encrypted data is sent to the cloud
- Value of prediction function is computed on encrypted data
- Encrypted result is sent back to the patient
- Patient enters key to decrypt answer.

- **Evaluation takes 0.2 seconds in the cloud!**



GUILLAUME PAUMIER/WIKIMEDI

Genetic gold. Each spot in a DNA microarray, such as this one, contains large amounts of sensitive genetic information.

How to Hide Your Genome

By Thomas Sumner | 16 February 2014 6:45 pm | 4 Comments

Thomas is a news intern at *Science*.

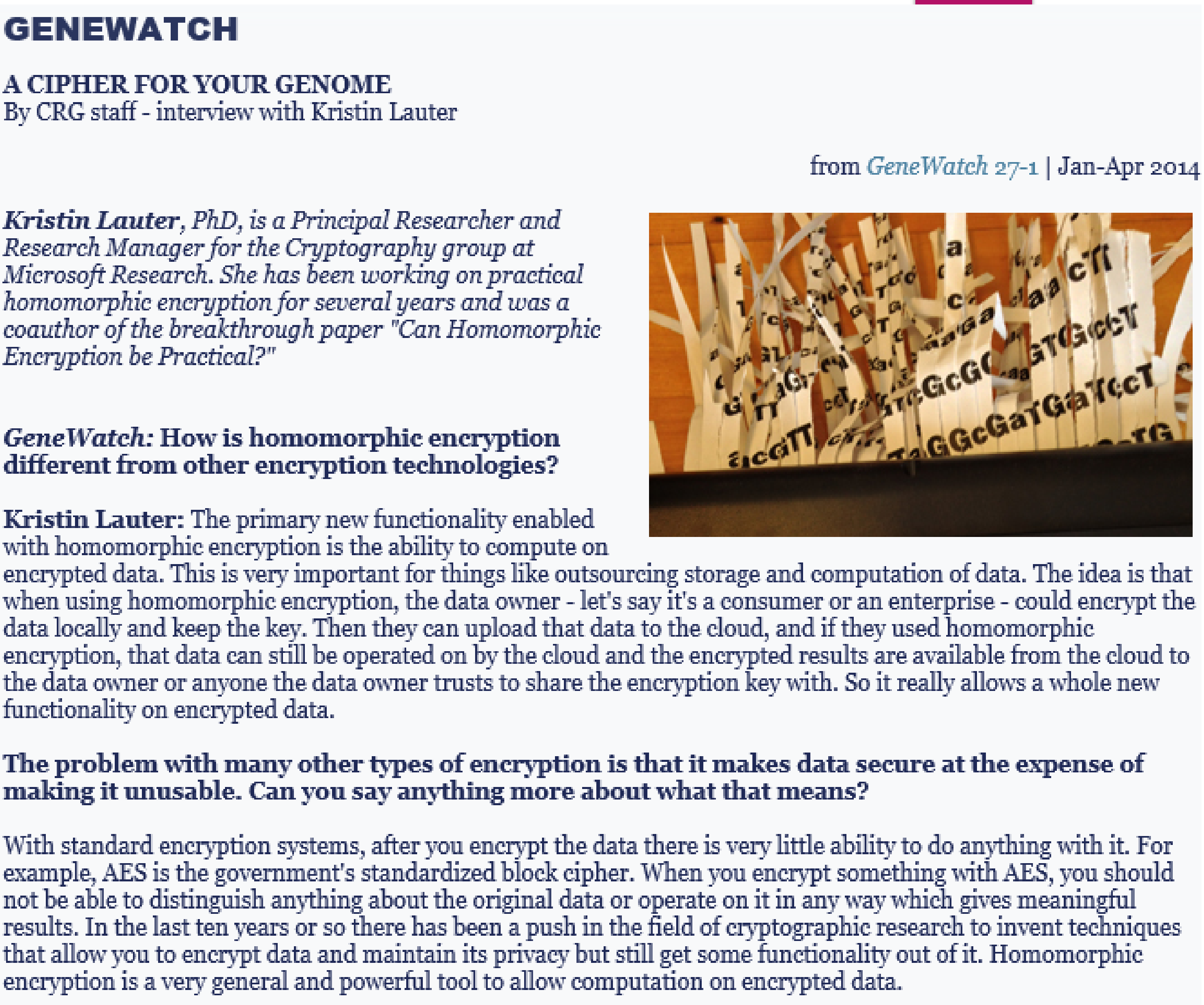
Email Thomas

Follow @SumnerSci

CHICAGO, ILLINOIS—As the cost of genetic sequencing plummets, experts believe our genomes will help doctors detect diseases and save lives. But not all of us are comfortable releasing our biological blueprints into the world. Now, cryptologists are perfecting a new privacy tool that turns genetic information into a secure yet functional format. Called homomorphic encryption and presented here today at the annual meeting of AAAS, which publishes *Science,* the method could help keep genomes private even as genetic testing shifts to cheap online cloud services.

Existing encryption techniques make data secure at the expense of making it unusable. Because of this, most genetic sequences are simply anonymized before being sent out for analysis. However, computational biologist Yaniv Erlich at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, told meeting attendees that with a little genetic sleuthing, this supposedly anonymous data can easily track back to its owner. Erlich says he positively matched 12% of male genomes with the exact person they originated from.

In 2009, the first lattice-based cryptography scheme was announced by IBM. The geometry-based encryption method allows data to be manipulated through both multiplication and addition while remaining encrypted. Researchers realized that the complex algorithms used during genetic tests could be closely approximated by the two basic mathematical operations. Lattice cryptology enabled homomorphic encryption, allowing computers to analyze encrypted data and return encrypted results without ever being able to decode the information. Cryptologist Kristin Lauter, research manager for the cryptography group at Microsoft Research in Redmond, Washington, likened the method to locking a gold brick in a



GENEWATCH

A CIPHER FOR YOUR GENOME

By CRG staff - interview with Kristin Lauter

from *GeneWatch* 27-1 | Jan-Apr 2014

***Kristin Lauter**, PhD, is a Principal Researcher and Research Manager for the Cryptography group at Microsoft Research. She has been working on practical homomorphic encryption for several years and was a coauthor of the breakthrough paper "Can Homomorphic Encryption be Practical?"*

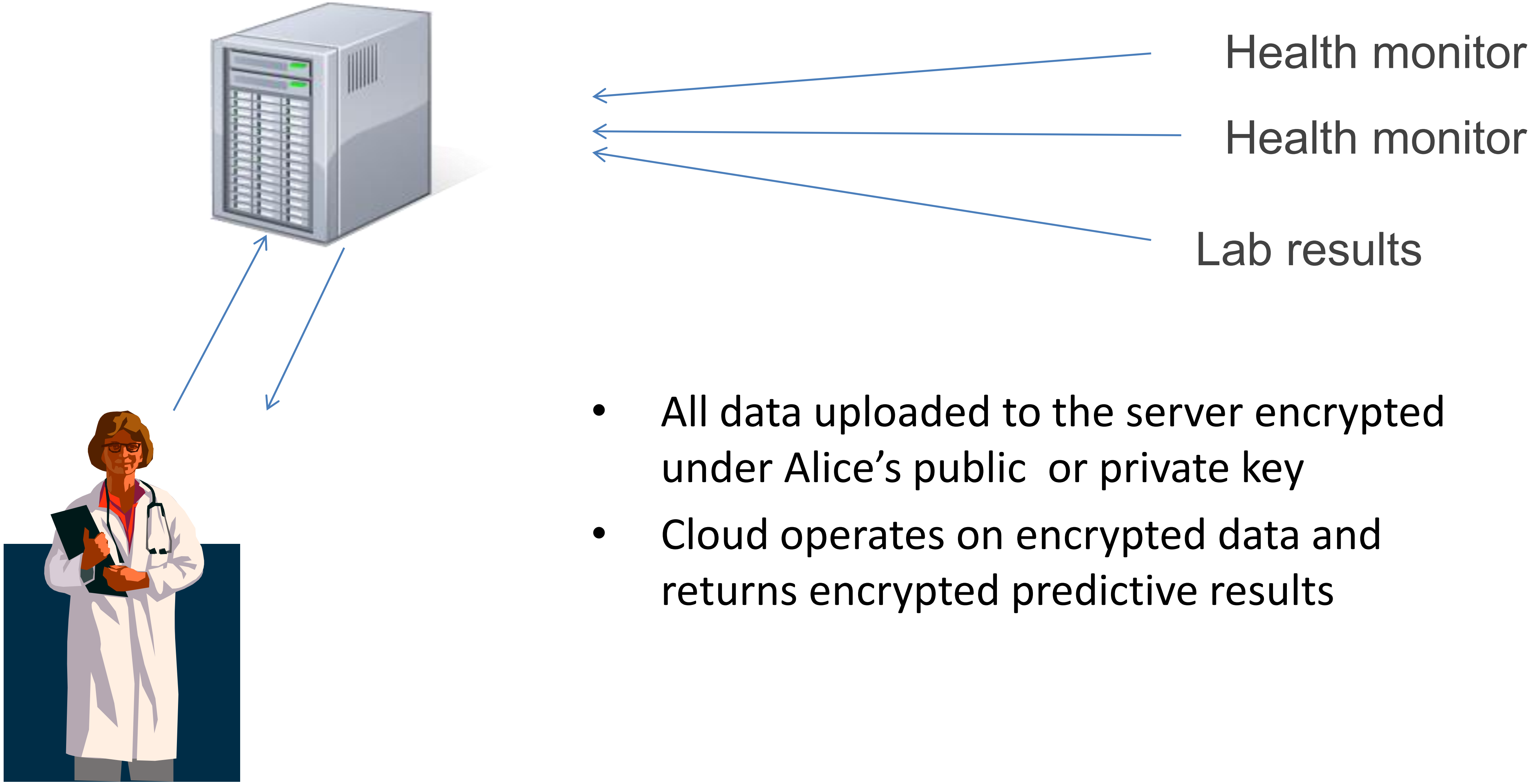
***GeneWatch:* How is homomorphic encryption different from other encryption technologies?**

Kristin Lauter: The primary new functionality enabled with homomorphic encryption is the ability to compute on encrypted data. This is very important for things like outsourcing storage and computation of data. The idea is that when using homomorphic encryption, the data owner - let's say it's a consumer or an enterprise - could encrypt the data locally and keep the key. Then they can upload that data to the cloud, and if they used homomorphic encryption, that data can still be operated on by the cloud and the encrypted results are available from the cloud to the data owner or anyone the data owner trusts to share the encryption key with. So it really allows a whole new functionality on encrypted data.

The problem with many other types of encryption is that it makes data secure at the expense of making it unusable. Can you say anything more about what that means?

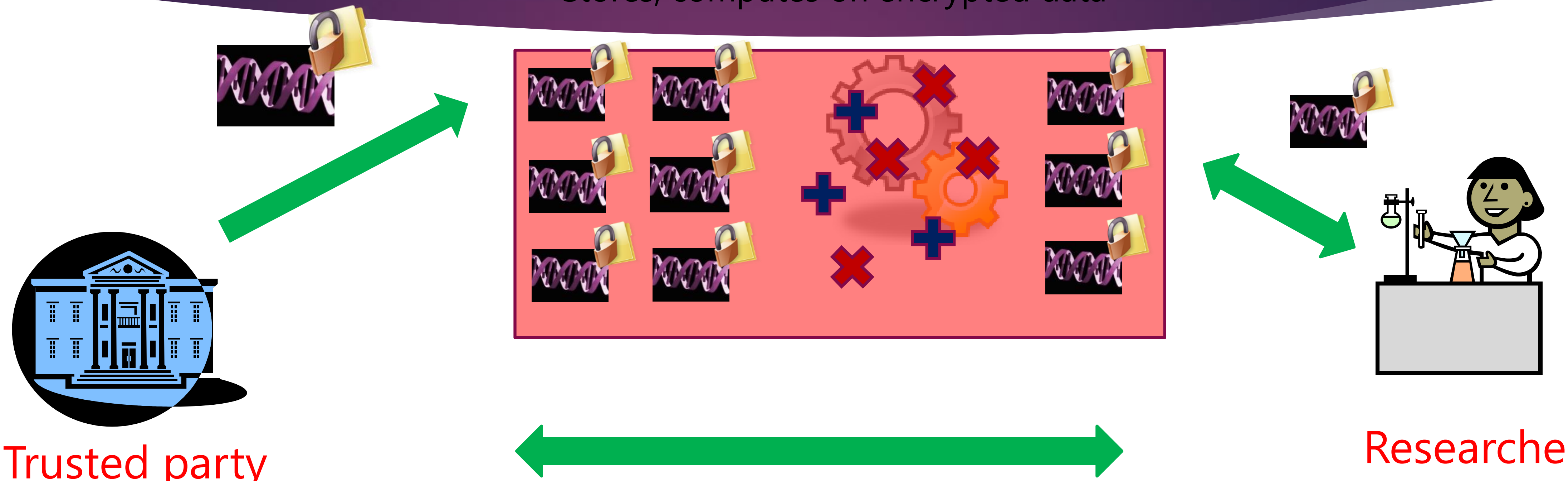
With standard encryption systems, after you encrypt the data there is very little ability to do anything with it. For example, AES is the government's standardized block cipher. When you encrypt something with AES, you should not be able to distinguish anything about the original data or operate on it in any way which gives meaningful results. In the last ten years or so there has been a push in the field of cryptographic research to invent techniques that allow you to encrypt data and maintain its privacy but still get some functionality out of it. Homomorphic encryption is a very general and powerful tool to allow computation on encrypted data.

Processing of encrypted medical data



- All data uploaded to the server encrypted under Alice’s public or private key
- Cloud operates on encrypted data and returns encrypted predictive results

Scenario for genomic data



Untrusted cloud service

Stores, computes on encrypted data

Trusted party

hosts data and
regulates access

Requests for decryption of results

(requires a policy)

Researcher:

requests encrypted
results of specific
computations

What kinds of computation?

- Building predictive models
- Predictive analysis
  - Classification tasks
  - Disease prediction
  - Sequence matching
- Data quality testing
- Basic statistical functions
- Statistical computations on genomic data

Functions to compute

- Average, Standard deviation, Chi-squared, …
- Logistical regression: the prediction is

$$f(x) = e^x/(1+e^x)$$

where x is the sum of $\alpha_i\ x_i$, where α_i is the weighting constant or regression coefficient for the variable x_i

Machine Learning for Predictive Modeling

Supervised Learning

Goal: derive a function from labeled training data

Outcome: use the “learned” function to give a prediction (label) on new data

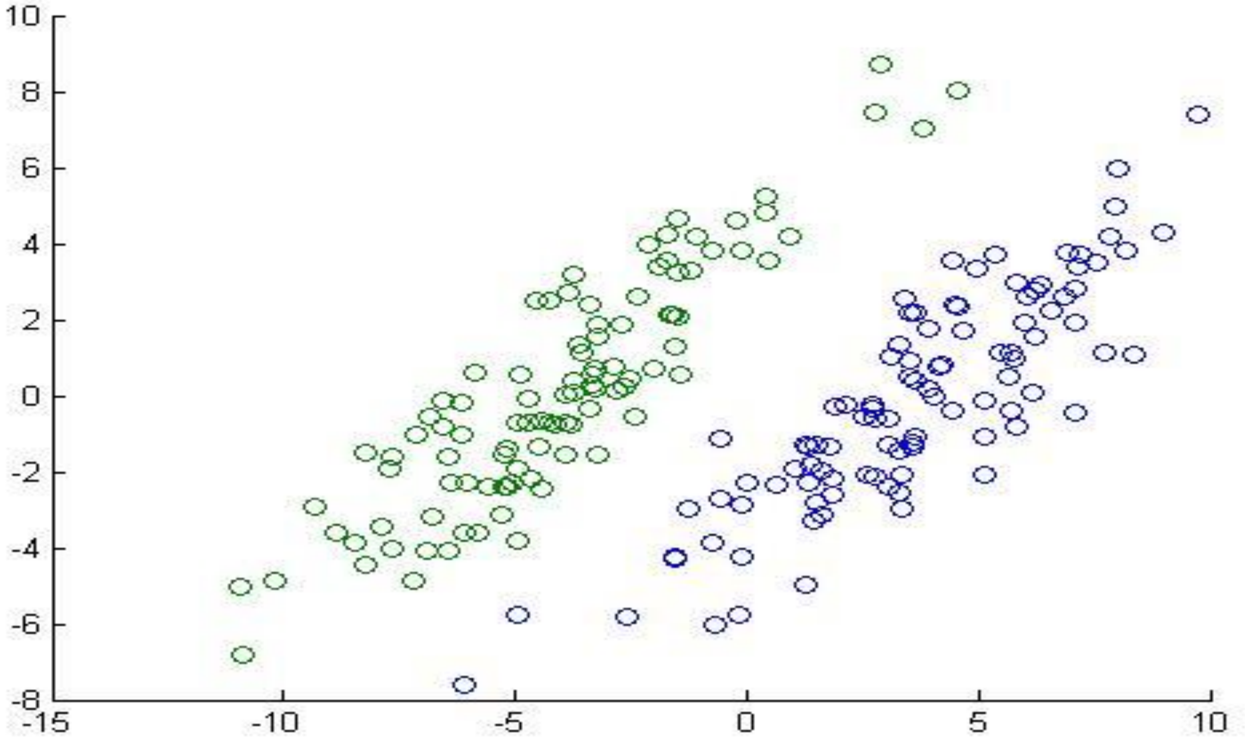
Training data represented as vectors.

Linear Means Classifier (binary)

- Divide training data into (two) classes according to their label
- Compute mean vectors for each class
- Compute difference between means
- Compute the midpoint
- Define a hyperplane between the means, separating the two classes

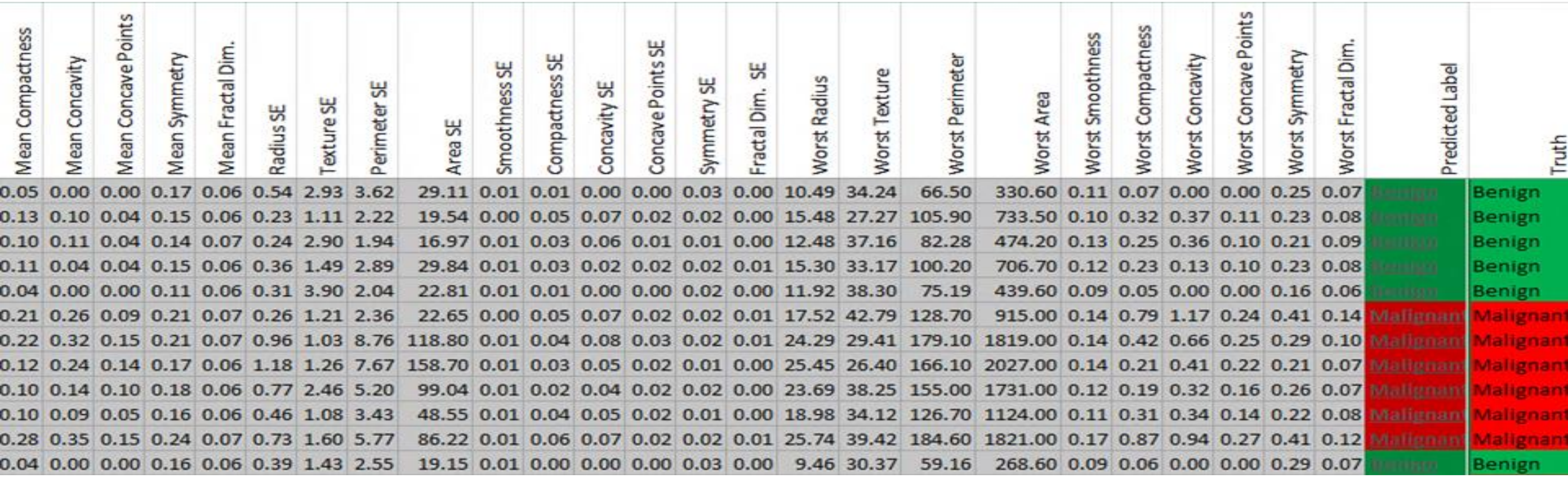
Binary classification example

- FD



Predictions on Medical data

Tumor measurements: Benign or Malignant



Mean Compactness	Mean Concavity	Mean Concave Points	Mean Symmetry	Mean Fractal Dim.	Radius SE	Texture SE	Perimeter SE	Area SE	Smoothness SE	Compactness SE	Concavity SE	Concave Points SE	Symmetry SE	Fractal Dim. SE	Worst Radius	Worst Texture	Worst Perimeter	Worst Area	Worst Smoothness	Worst Compactness	Worst Concavity	Worst Concave Points	Worst Symmetry	Worst Fractal Dim.	Predicted Label	Truth
0.05	0.00	0.00	0.17	0.06	0.54	2.93	3.62	29.11	0.01	0.01	0.00	0.00	0.03	0.00	10.49	34.24	66.50	330.60	0.11	0.07	0.00	0.00	0.25	0.07	Benign	Benign
0.13	0.10	0.04	0.15	0.06	0.23	1.11	2.22	19.54	0.00	0.05	0.07	0.02	0.02	0.00	15.48	27.27	105.90	733.50	0.10	0.32	0.37	0.11	0.23	0.08	Benign	Benign
0.10	0.11	0.04	0.14	0.07	0.24	2.90	1.94	16.97	0.01	0.03	0.06	0.01	0.01	0.00	12.48	37.16	82.28	474.20	0.13	0.25	0.36	0.10	0.21	0.09	Benign	Benign
0.11	0.04	0.04	0.15	0.06	0.36	1.49	2.89	29.84	0.01	0.03	0.02	0.02	0.02	0.01	15.30	33.17	100.20	706.70	0.12	0.23	0.13	0.10	0.23	0.08	Benign	Benign
0.04	0.00	0.00	0.11	0.06	0.31	3.90	2.04	22.81	0.01	0.01	0.00	0.00	0.02	0.00	11.92	38.30	75.19	439.60	0.09	0.05	0.00	0.00	0.16	0.06	Benign	Benign
0.21	0.26	0.09	0.21	0.07	0.26	1.21	2.36	22.65	0.00	0.05	0.07	0.02	0.02	0.01	17.52	42.79	128.70	915.00	0.14	0.79	1.17	0.24	0.41	0.14	Malignant	Malignant
0.22	0.32	0.15	0.21	0.07	0.96	1.03	8.76	118.80	0.01	0.04	0.08	0.03	0.02	0.01	24.29	29.41	179.10	1819.00	0.14	0.42	0.66	0.25	0.29	0.10	Malignant	Malignant
0.12	0.24	0.14	0.17	0.06	1.18	1.26	7.67	158.70	0.01	0.03	0.05	0.02	0.01	0.00	25.45	26.40	166.10	2027.00	0.14	0.21	0.41	0.22	0.21	0.07	Malignant	Malignant
0.10	0.14	0.10	0.18	0.06	0.77	2.46	5.20	99.04	0.01	0.02	0.04	0.02	0.02	0.00	23.69	38.25	155.00	1731.00	0.12	0.19	0.32	0.16	0.26	0.07	Malignant	Malignant
0.10	0.09	0.05	0.16	0.06	0.46	1.08	3.43	48.55	0.01	0.04	0.05	0.02	0.01	0.00	18.98	34.12	126.70	1124.00	0.11	0.31	0.34	0.14	0.22	0.08	Malignant	Malignant
0.28	0.35	0.15	0.24	0.07	0.73	1.60	5.77	86.22	0.01	0.06	0.07	0.02	0.02	0.01	25.74	39.42	184.60	1821.00	0.17	0.87	0.94	0.27	0.41	0.12	Malignant	Malignant
0.04	0.00	0.00	0.16	0.06	0.39	1.43	2.55	19.15	0.01	0.00	0.00	0.00	0.03	0.00	9.46	30.37	59.16	268.60	0.09	0.06	0.00	0.00	0.29	0.07	Benign	Benign

Machine Learning on Encrypted Data

- Implements Polynomial Machine Learning Algorithms
- Integer Algorithms
- Division-Free Linear Means Classifier (DFI-LM)
- Fisher’s Linear Discriminant Classifier

DFI-LM experiments

$(P_1)\ q = 2^{128},\ t = 2^{15},\ \sigma = 16,\ d = 4096$

SH.Keygen	SH.Enc	SH.Dec(2)	SH.Dec(3)	SH.Add	SH.Mult
156	379	29	52	1	106

Timing in ms in Magma on a single core of an Intel Core i5 CPU650 @ 3.2 GHz. 128-bit security with distinguishing advantage 2^{-64}.

data	# features	algorithm	train	classify
surrogate	2	linear means	230	235
Iris	4	linear means	510	496

Statistics on Genomic Data

- **Pearson Goodness-Of-Fit Test**
  - checks data for bias (Hardy-Weinberg equilibrium)
- **Cochran-Armitage Test for Trend**
  - Determine **correlation** between genome and traits
- **Linkage Disequilibrium Statistic**
  - Estimates correlations between genes
  - **Estimation Maximization (EM) algorithm for haplotyping**

Genomic algorithm performance

80-bit security

- Parameter set I: $n = 4096,\ q \approx 2^{192}$, ciphertext $\approx 100\text{KB}$
- Parameter set II: $n = 8192,\ q \approx 2^{384}$, ciphertext $\approx 400\text{KB}$

Algorithm	Pearson	EM (iterations)			LD	CATT
		1	2	3		
Parameters I	0.3s	0.6s	1.1s	-	0.2s	1.0s
Parameters II	1.4s	2.3s	4.5s	6.9s	0.7s	3.6s

Proof-of-concept implementation: computer algebra system Magma,
Intel Core i7 @ 3.1GHz, 64-bit Windows 8.1

Performance Summary

- Data quality (Pearson Goodness-of-Fit)

~ 0.3 seconds, 1,000 patients

- Predicting Heart Attack (Logistic Regression)

~ 0.2 seconds

- Building models (Linear Means Classifier)

~0.9 secs train, classify: 30 features, 100 training samples

- Sequence matching (Edit distance)

~27 seconds amortized, length 8

Core i7 3.4GHz

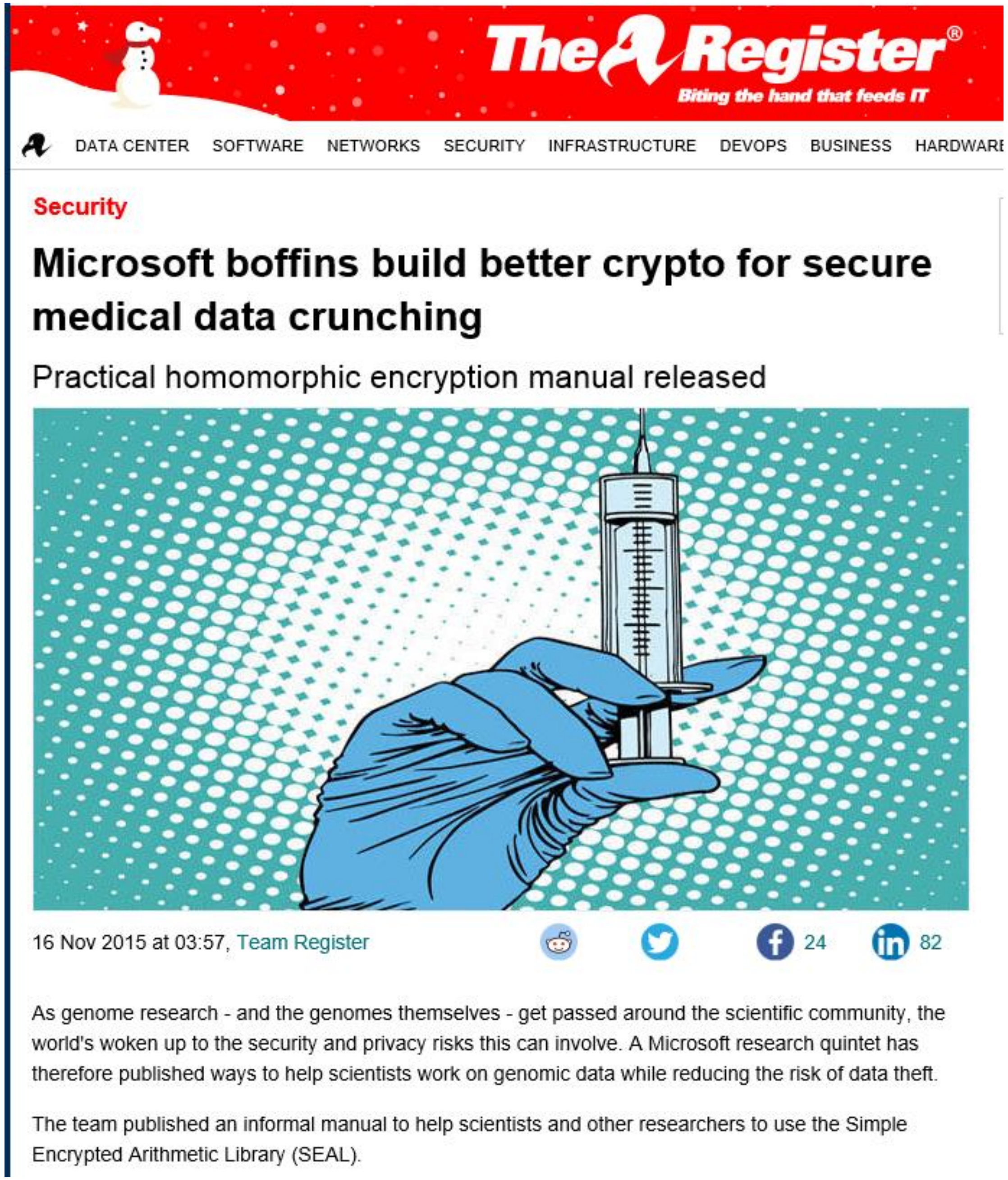
80-bit security

Practical Homomorphic Encryption

- do not need *fully* homomorphic encryption
- encode integer information as "integers"
- several orders of magnitude speed-up
- do not need deep circuits to do a single multiplication
- for "logical" circuits, use ciphertext packing and tradeoff depth for ciphertext size
- need to set parameters to ensure correctness and security
- PHE=homomorphic for any fixed circuit size, with correctly chosen parameters

SEAL: Simple Encrypted Arithmetic Library

- SEAL public release in 2015 by Microsoft Research for research purposes
- Freely, publicly available
- Compare to HELib from IBM (~2013)
- Includes automatic parameter selection for user-defined tasks
- A rush of unexpected press coverage in popular media



The Register®
Biting the hand that feeds IT

Security

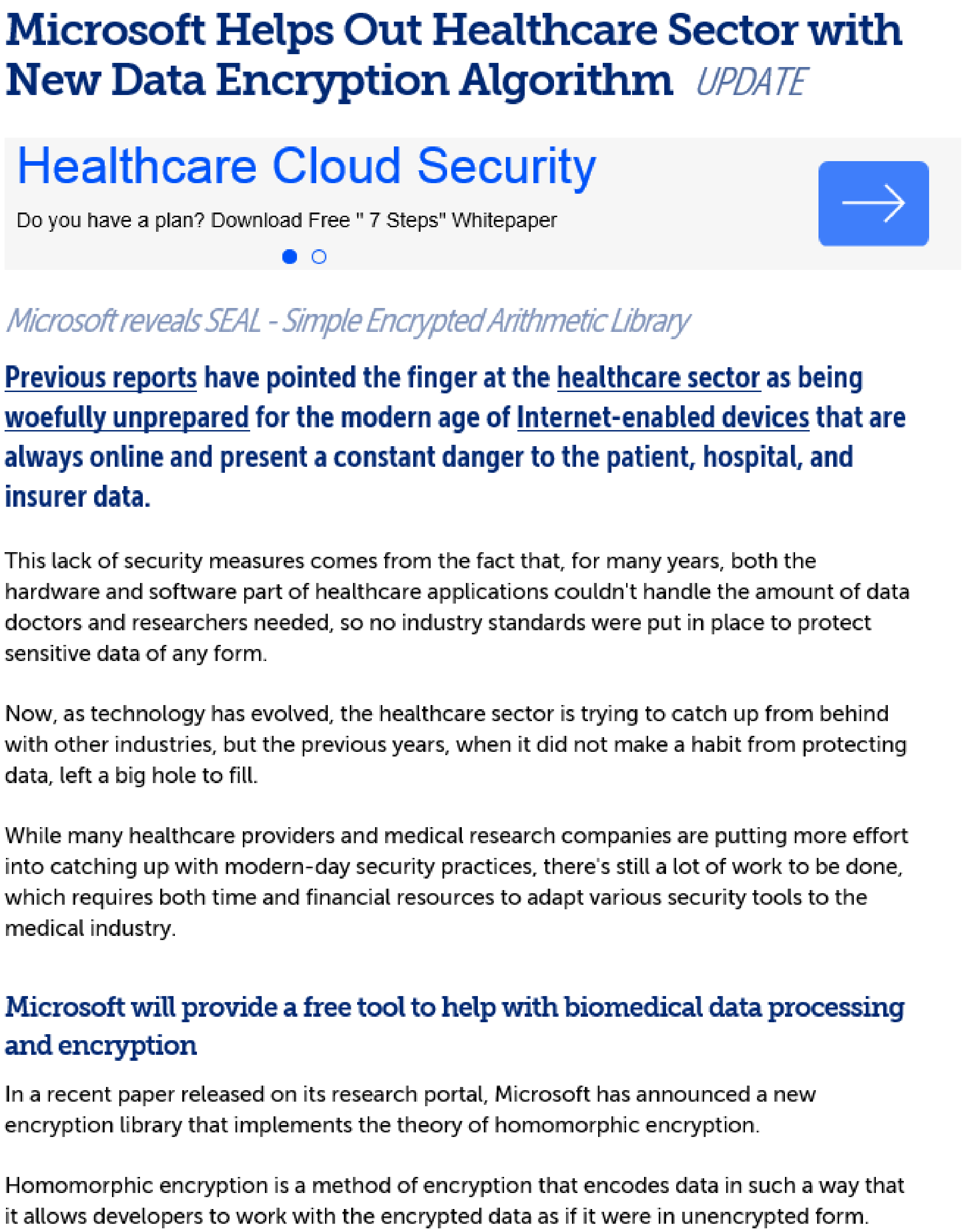
Microsoft boffins build better crypto for secure medical data crunching

Practical homomorphic encryption manual released

16 Nov 2015 at 03:57, Team Register

As genome research - and the genomes themselves - get passed around the scientific community, the world's woken up to the security and privacy risks this can involve. A Microsoft research quintet has therefore published ways to help scientists work on genomic data while reducing the risk of data theft.

The team published an informal manual to help scientists and other researchers to use the Simple Encrypted Arithmetic Library (SEAL).



Microsoft Helps Out Healthcare Sector with New Data Encryption Algorithm *UPDATE*

Microsoft reveals SEAL - Simple Encrypted Arithmetic Library

Previous reports have pointed the finger at the healthcare sector as being woefully unprepared for the modern age of Internet-enabled devices that are always online and present a constant danger to the patient, hospital, and insurer data.

This lack of security measures comes from the fact that, for many years, both the hardware and software part of healthcare applications couldn't handle the amount of data doctors and researchers needed, so no industry standards were put in place to protect sensitive data of any form.

Now, as technology has evolved, the healthcare sector is trying to catch up from behind with other industries, but the previous years, when it did not make a habit from protecting data, left a big hole to fill.

While many healthcare providers and medical research companies are putting more effort into catching up with modern-day security practices, there's still a lot of work to be done, which requires both time and financial resources to adapt various security tools to the medical industry.

Microsoft will provide a free tool to help with biomedical data processing and encryption

In a recent paper released on its research portal, Microsoft has announced a new encryption library that implements the theory of homomorphic encryption.

Homomorphic encryption is a method of encryption that encodes data in such a way that it allows developers to work with the encrypted data as if it were in unencrypted form.

What are the Costs? Challenges? Obstacles?

For homomorphic encryption

- Storage costs (large ciphertexts)
- New hard problems (introduced 2010-2015)
- Efficiency at scale (large amounts of data, deep circuits)

For Garbled Circuits

- High interaction costs
- Bandwidth use
- Integrate with storage solutions

Challenges for the future:

- Public Databases: multiple patients under different keys
- More efficient encryption at scale
- Integrate with other crypto solutions
- Expand functionality
- Attack underlying hard problems

High-level message

- Importance of mathematics for solving current, real-world problems for society
  - New SIAM Journal SIAAG on Applications of Algebra and Geometry
- Positive impact of applications on mathematics
- Engaging across disciplines creates a fun and productive research environment
- Examples of engaging with the public through the popular press

Joint work with:

…and thanks to iDASH and co-authors for selected slides…

- SEAL Team: Kim Laine, John Wernsing, Michael Naehrig, Ran Gilad-Bachrach, Nathan Dowlin, Kristin Lauter
- Can Homomorphic Encryption be Practical?
  - Kristin Lauter, Michael Naehrig, Vinod Vaikuntanathan, CCSW 2011
- ML Confidential: Machine Learning on Encrypted Data
  - Thore Graepel, Kristin Lauter, Michael Naehrig, ICISC 2012
- Predictive Analysis on Encrypted Medical Data
  - Joppe W. Bos, Kristin Lauter, and Michael Naehrig, Journal of Biomedical Informatics, 2014.
- Private Computation on Encrypted Genomic Data
  - Kristin Lauter, Adriana Lopez-Alt, Michael Naehrig, GenoPri2014, LatinCrypt2014.
- Homomorphic Computation of Edit Distance
  - Jung Hee Cheon, Miran Kim, Kristin Lauter, WAHC, FC 2015
- RLWE Attacks:
  - Kirsten Eisentraeger, Sean Hallgren, Kate Stange, Ekin Ozman, Yara Elias, Hao Chen, SAC '14, Crypto '15