DIY DNA OSINT!

or... how to enhance your social engineering skills
using recent genomics

Min2RIEN - CNRS, FRANCE – November, 19 2019

Renaud Lifchitz

digital.security

Renaud Lifchitz:

speaker's bio

- French senior security engineer

- Main activities:
  - Penetration testing & security audits
  - Security research
  - Security trainings

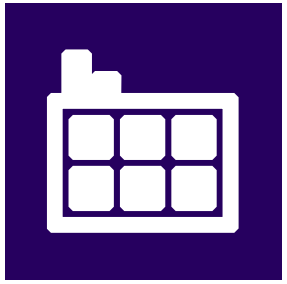
- Main interests:
  - IoT security (hardware & RF)
  - Security of protocols (authentication, cryptography, information leakage, reverse engineering...)
  - Secure programming
  - Number theory (integer factorization, primality testing, ...)

digital.security core business is based on **3 fields of expertise** driven by **6 services**

EXPERTISE

Information Systems Security

Internet of Things Security



Industrial IT Security

SERVICES

Audit

Consulting

Training

BUG BOUNTY MANAGEMENT

CERT Services

Operational Security

Integration & Projects

(*) PASSI every domains, PASSI LPM, PASSI Monaco

Our Lab & CERT

The Laboratory of Digital Security is a technology sanctuary in which we conduct digital investigations and all types of analyzes on smart devices and their ecosystem to detect the smallest vulnerabilities.

This laboratory allows us to deliver an IoT security label which guarantees a security level compliance with the requirements.

Analysis and research on radio frequency protocols and detection of radiating equipment

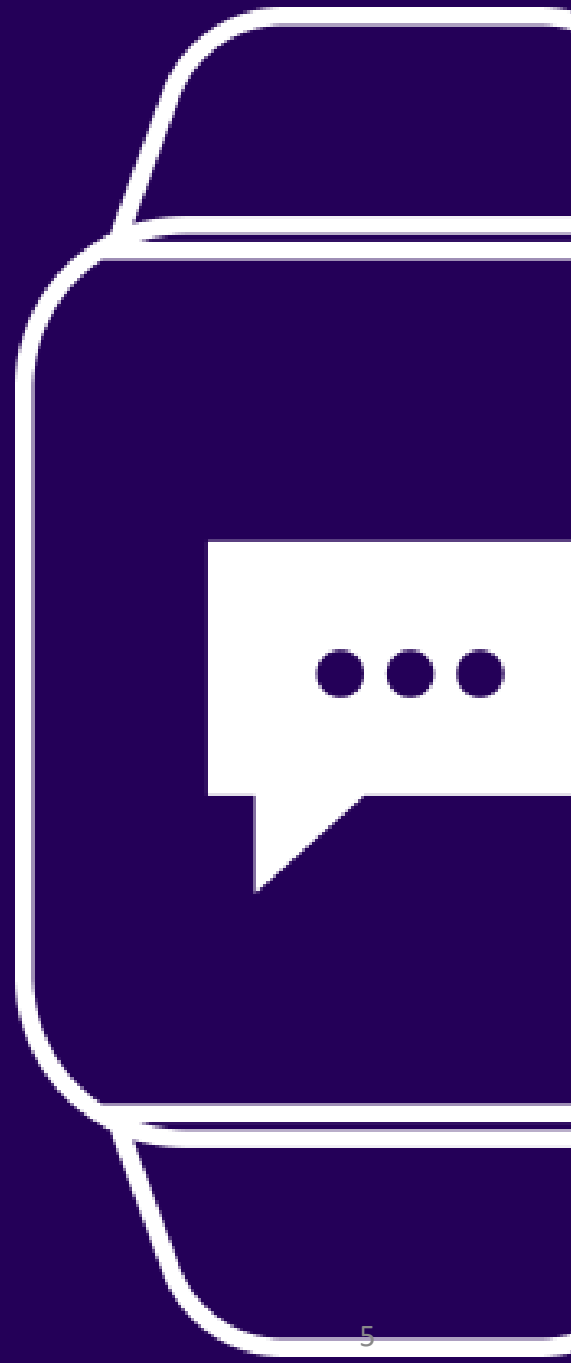
"Physical" attacks and tests on smart devices

Forensics to identify and secure digital evidences

INTRO

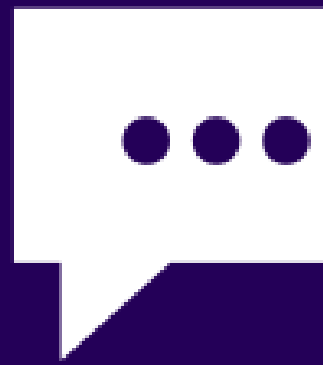
The first whole human genome sequencing took years of effort and cost about 2.2 billion euros in 2003. Today it takes a few weeks and a few hundreds euros to get your own (or someone else's!) DNA sequenced. **More than 20 millions of US citizens have already sequenced their DNA** and several hundreds of raw DNA files are available through the Web, sometimes without their owner's consent... Even if DNA is not a completely documented format, many things can be found against people with their DNA available.

What are today's tools to study DNA? Are they freely available? What can you really find about somebody? How is it related to information security? Learn many things about you in this first of its kind talk!



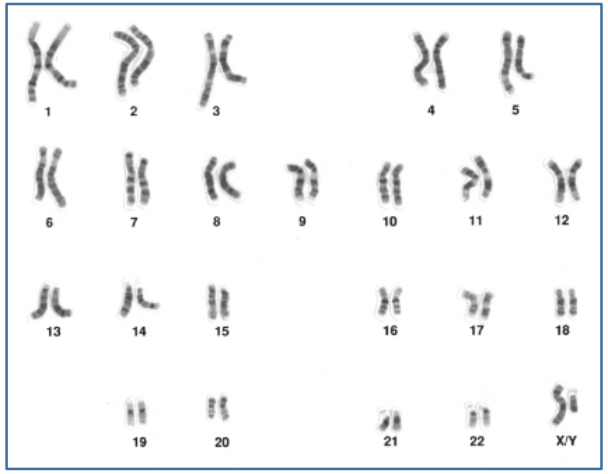
OUTLINE

- Generalities about DNA and genes
- Sequencing services: differences between the bad & the good
- DNA file formats
- Open source genes databases & tools
- Interesting online services
- DNA OSINT sources
- How to find nearly anybody with DNA
- Find your opponent's strengths and weaknesses using DNA
- Recommendations about DNA and privacy



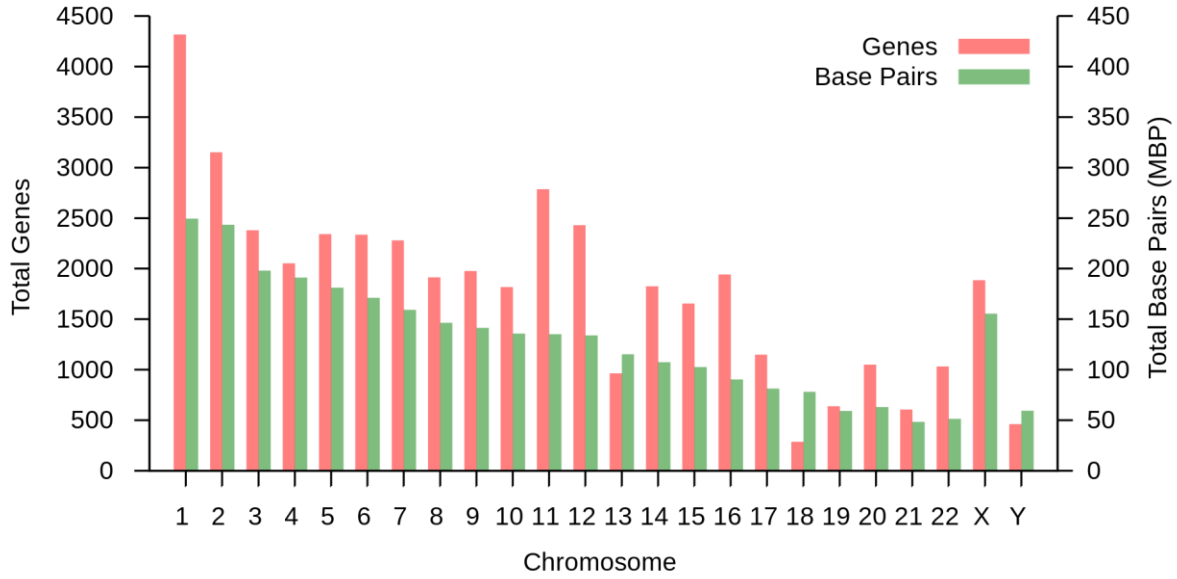
Generalities about DNA and genes (1/4)

- Humans:
  - 46 chromosomes:
    22 autosomal pairs + 2 sex chromosomes
  - DNA structure found in 1953
  - First human genome sequencing finished only in 2003
  - Everyone has 2 sequences of DNA
  - A genotype has 2 alleles



Generalities about DNA and genes (2/4)

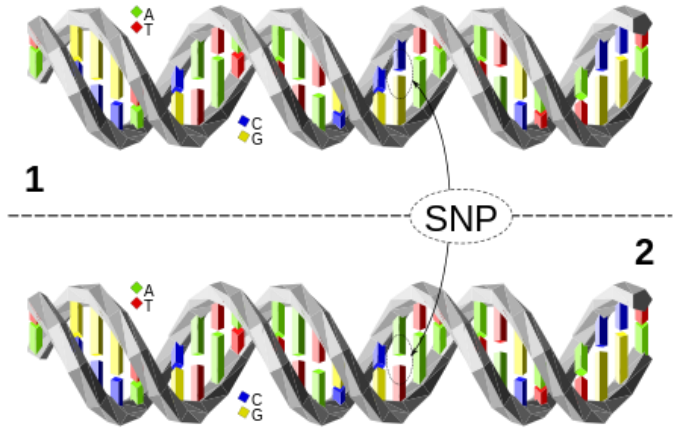
- Human DNA:
  - 2 strands of only 4 kinds of molecules : A, C, G and T
  - 3 billion base pairs (nucleotides)
  - about 30,000 genes



https://en.wikipedia.org/wiki/Human_genome

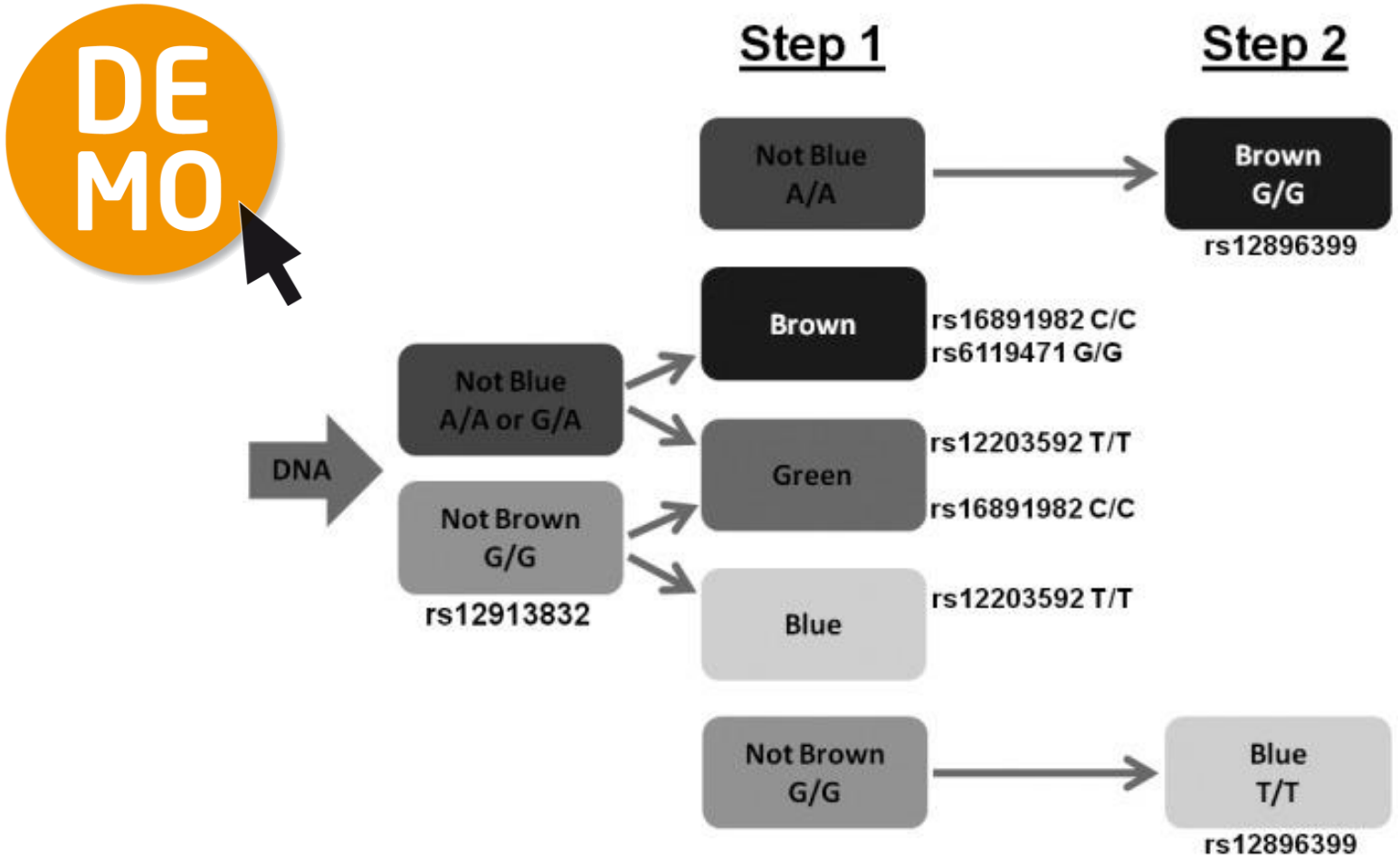
Generalities about DNA and genes (3/4)
Human SNPs

- Humans nucleotides are 99,9% similar
- 90% of variations affect a single nucleotide: SNP (Single Nucleotide Polymorphism)
- All human SNPs are known
- Example, rs16891982 SNP:
  - C/C genotype: dark hair, comes from Asia
  - G/G genotype: lighter hair, comes from Europe
  - C/G or G/C: medium hair
- A typical SNP can cause between 1.1 to 1.3 fold increase in «risk» - OR (Odds Ratio)



https://en.wikipedia.org/wiki/Single-nucleotide_polymorphism

Generalities about DNA and genes (4/4)
Example: eye color prediction



https://www.researchgate.net/publication/239525268_Improved_eye-_and_skin-color_prediction_based_on_8_SNPs

Sequencing services: differences between the bad and the good

- Now it's technically easy and affordable to sequence any DNA using DTC (direct-to-consumer) testing
- Illegal in some countries (France)
- Saliva spit or rubbed cheek

- Two kinds of sequencing:
  - Incomplete: microarray technology (measures known variability), from $50 to $150
  - Complete: WGS (Whole Genome Sequencing), from $300

- Results come a few weeks later!

DNA file formats

- Whole Genome Sequencing:
  - typical workflow:
    FASTQ ⇨ BAM ⇨ gVCF or VCF
- Microarray:
  - VCF (SNP differences with the reference genome)
  - 23andme
- At the end, one line for each SNP, around a million lines

DNA file formats
Google dorks: find DNA leaks!

- 23andme files (maybe in ZIP format):
  - `filetype:txt "rsid chromosome position genotype"`
  - `filetype:txt "rs16891982"`
  - `filetype:txt "23andMe" "rsid" "genotype"`
- gVCF & VCF files (maybe in gzip format):
  - `filetype:vcf "fileformat" "CHROM POS ID"`
  - `filetype:txt "fileformat" "CHROM POS ID"`
- Tip: Be sure to include all Google omitted results (end of page)

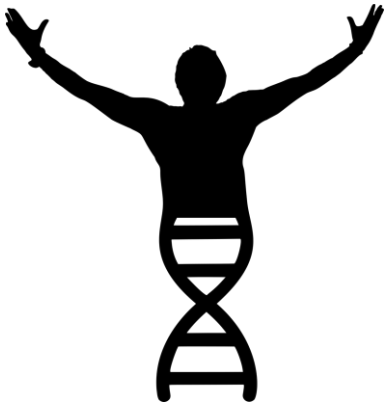
Open source genes databases & tools

- Genes for Good: https://genesforgood.sph.umich.edu/
- IGSR (The International Genome Sample Resource): http://www.internationalgenome.org/
- dbSNP: https://www.ncbi.nlm.nih.gov/snp/
- SNPedia: https://www.snpedia.com/index.php/SNPedia

How to find nearly anybody with DNA! (1/2)

- Ancestry:
  - Fatherhood: long sequences on the Y chromosome
  - Motherhood: long sequences on the mitochondrial DNA
- More than 20 millions of US citizens have already sequenced their DNA
- 1% of people sequenced would be enough to find anyone!

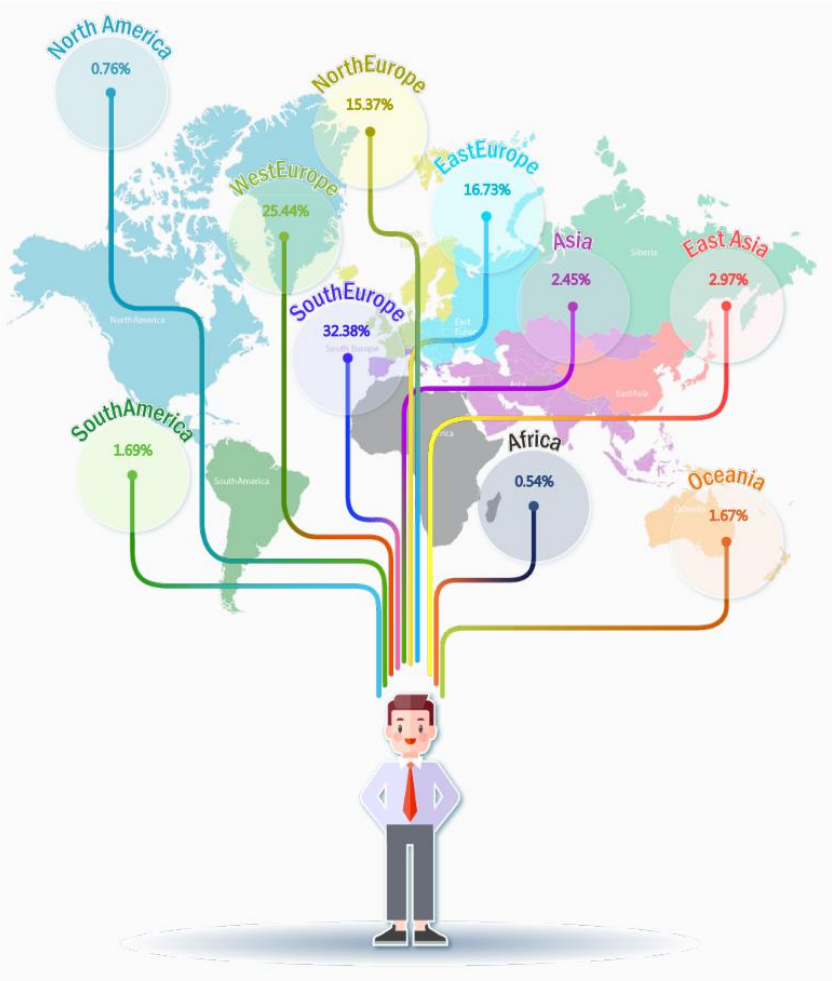
- Some DTC services to find relatives:
  - 23andme
  - Ancestry.com
  - MyHeritage
  - GEDMatch
  - Family Tree DNA
  - DNA.LAND
- Additionally, a few requests on Facebook would mostly complete the search!



How to find nearly anybody with DNA! (2/2)

Beware of ancestry services that pretend to give you your geographic origins, results can be very different depending:

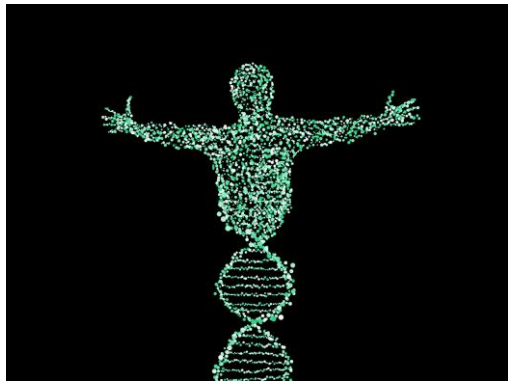
- on their customer base
- on their predefined regions
- on their methodology



Find your opponent's strengths & weaknesses using DNA

Lots of things can be found using a person's DNA:

- physical traits
- diseases
- allergies
- food preferences
- abilities
- weaknesses
- and even... personality traits!

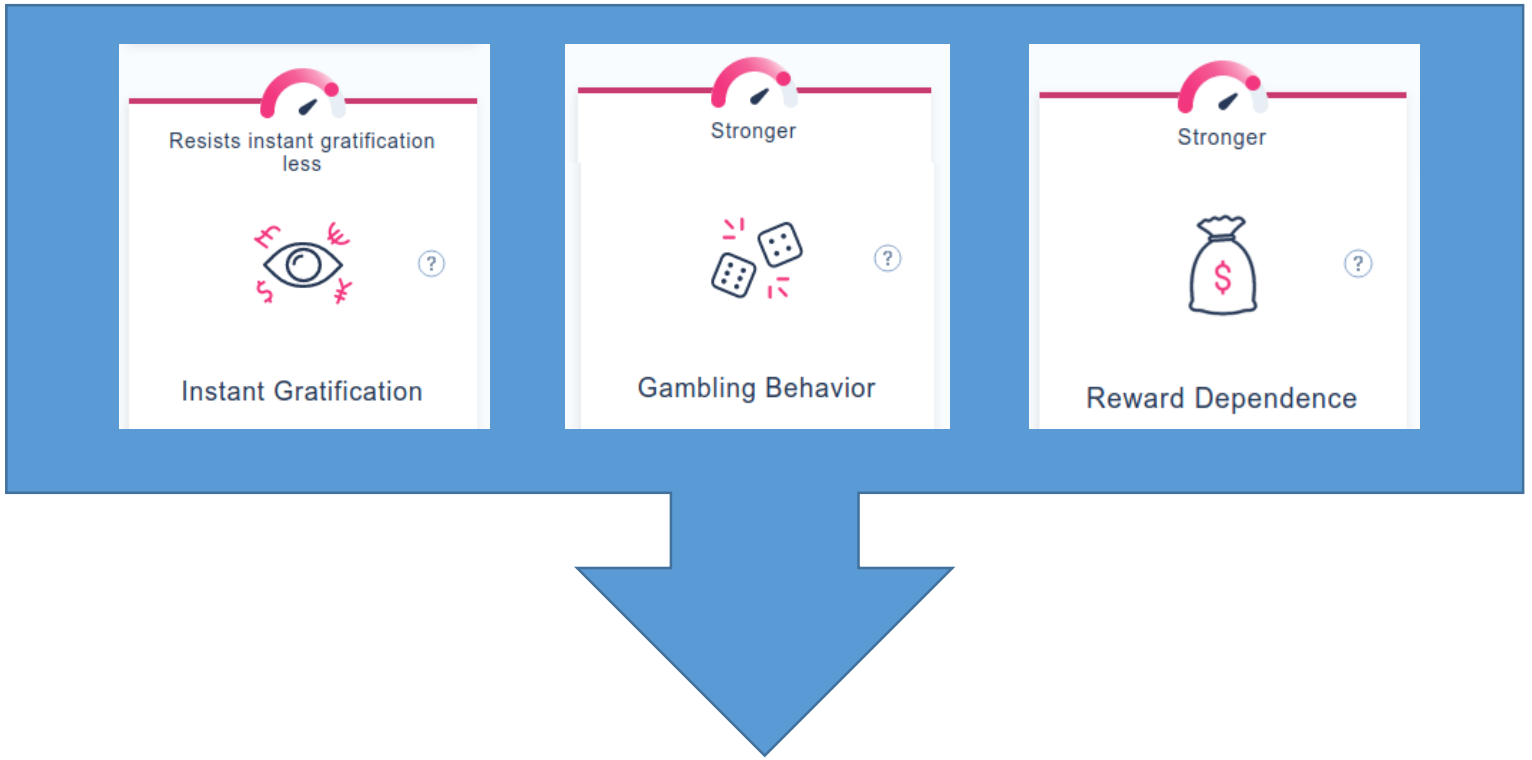


- Very useful for social engineering attacks...

DNA traits OSINT sources

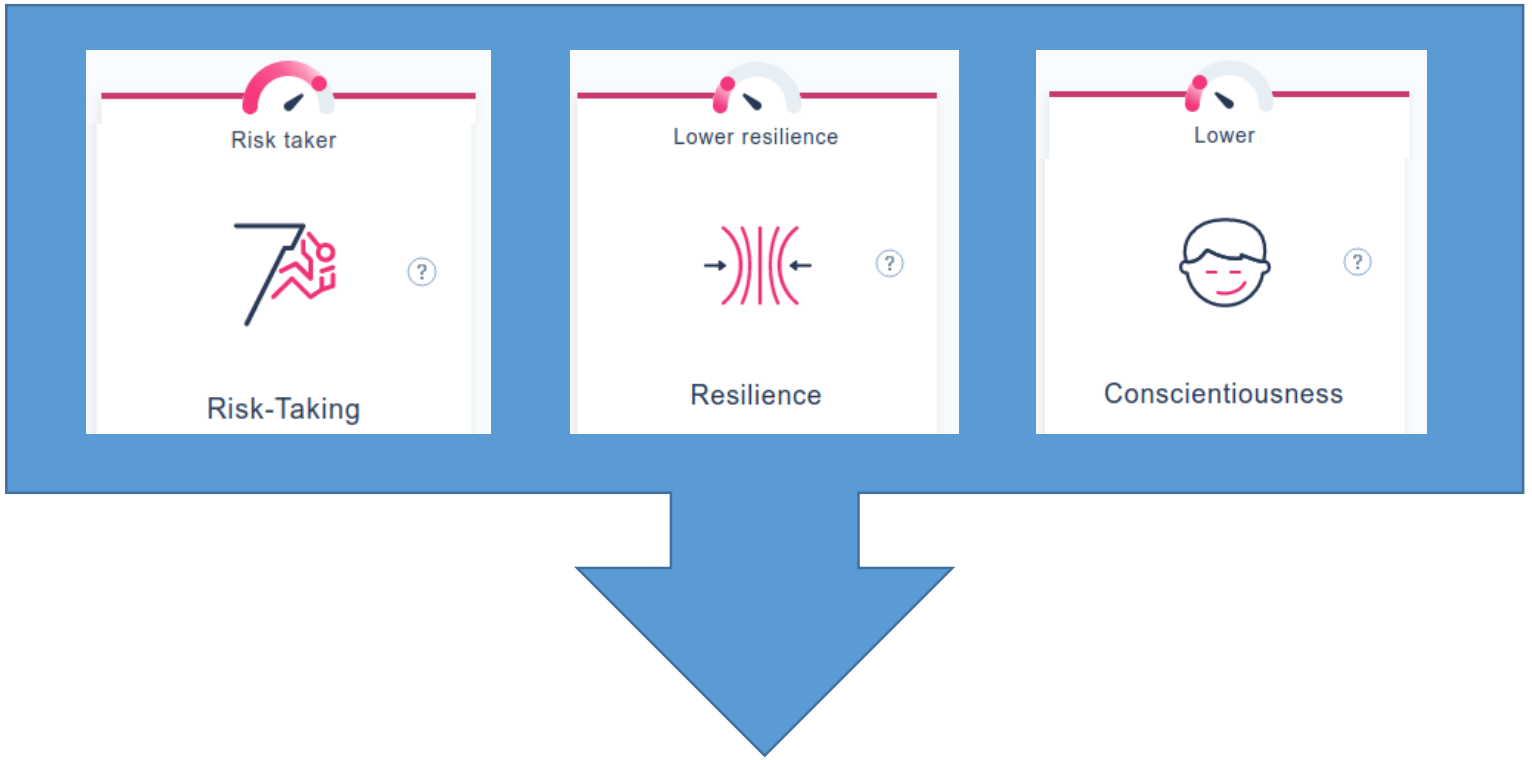
- Genomelink: https://genomelink.io/
- Promethease (cheap): https://promethease.com/
- Sequencing apps: https://sequencing.com/apps/app-market
- SelfDecode: https://www.selfdecode.com/
- Impute.me (free): https://www.impute.me/imputeme/

Find your opponent's strengths & weaknesses using DNA
Example 1



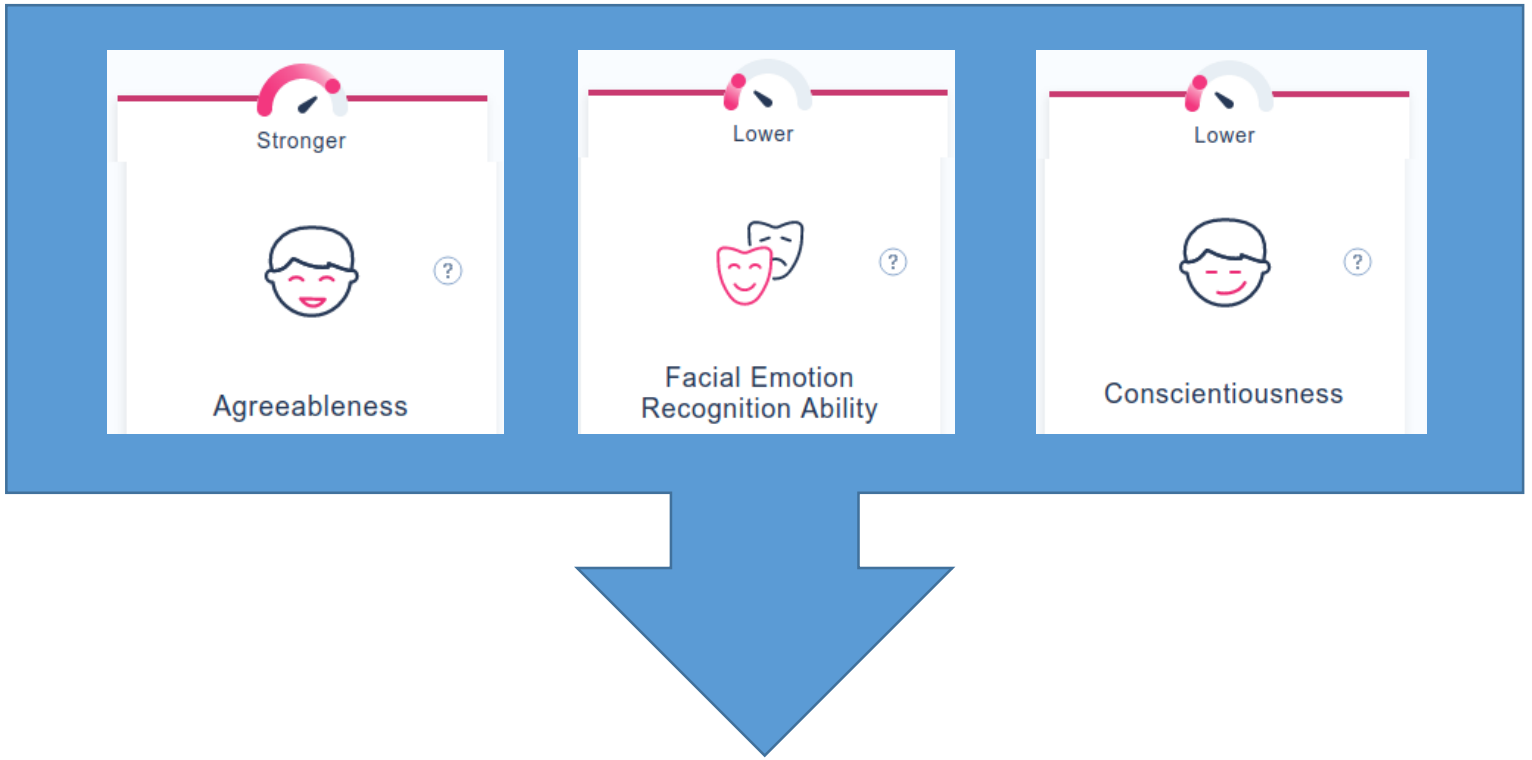
Easier spear-phishing using instant draw game!

Find your opponent's strengths & weaknesses using DNA
Example 2



Easier vishing attack!

Find your opponent's strengths & weaknesses using DNA
Example 3



Easier physical intrusion using impersonation!

Find your opponent's strengths & weaknesses using DNA Limitations

- Common traits are often based on several SNPs (ex.: more than 22 SNPs determine hair color - Eriksson et al., 2010)

- Very often non 100% deterministic:
  - Heritability (how much of a trait is explained by genetics)
  - OR («Odds Ratio»)
  - Also depends on environment & education, habits (epigenetics)

- Prefer results based on many studies, larger studies, and newer studies

Recommendations about DNA services & privacy

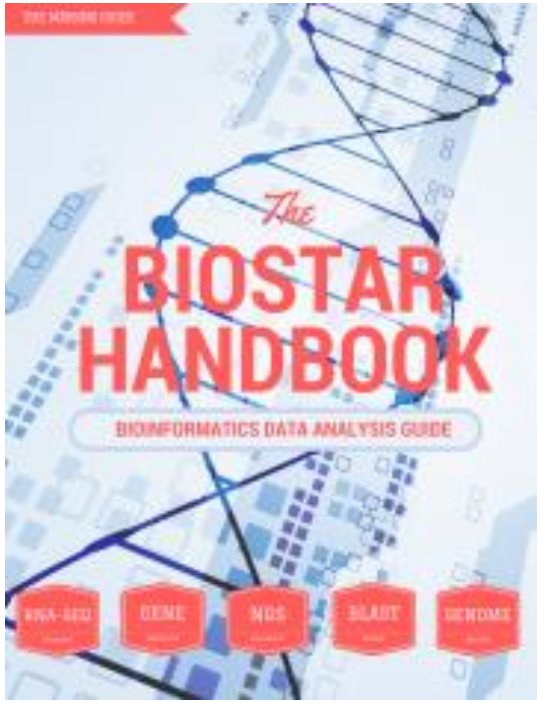
- Carefully read the service terms

- Compliance:
  - HIPAA
  - ISO27001
  - GDPR
    (General Data Protection Regulation)

- Prefer services with:
  - no selling & no sharing policy
  - actual security audits
  - «erase your data anytime» feature

Bibliography

- «Understand your DNA, a guide», Lasse Folkersen, 2019
- «The Family Tree guide to DNA Testing and Genetic Genealogy», April 2016, Blaine T. Bettinger
- «The Biostar Handbook: 2nd Edition», June 2019, www.biostarhandbook.com



Thank you!

Questions?

renaud.lifchitz@digital.security
Twitter: @nono2357

digital.security

Contact

info@digital.security

+33 (0)1 70 83 85 85

https://www.digital.security

Follow us

@Digital Security - Econocom

@iotcert

Offices

50 avenue Daumesnil
Immeuble B
75012 Paris
FRANCE

76 route de la demi lune, immeuble Madeleine
92057 Paris
FRANCE

13 bis Avenue Albert Einstein,
69100 Villeurbanne
FRANCE

Bastion Tower
5 Place du Champ de Mars
1050 Bruxelles
BELGIUM

144 rue Scheleck
L-3225 Bettembourg
LUXEMBOURG