Evolutions in molecular techniques

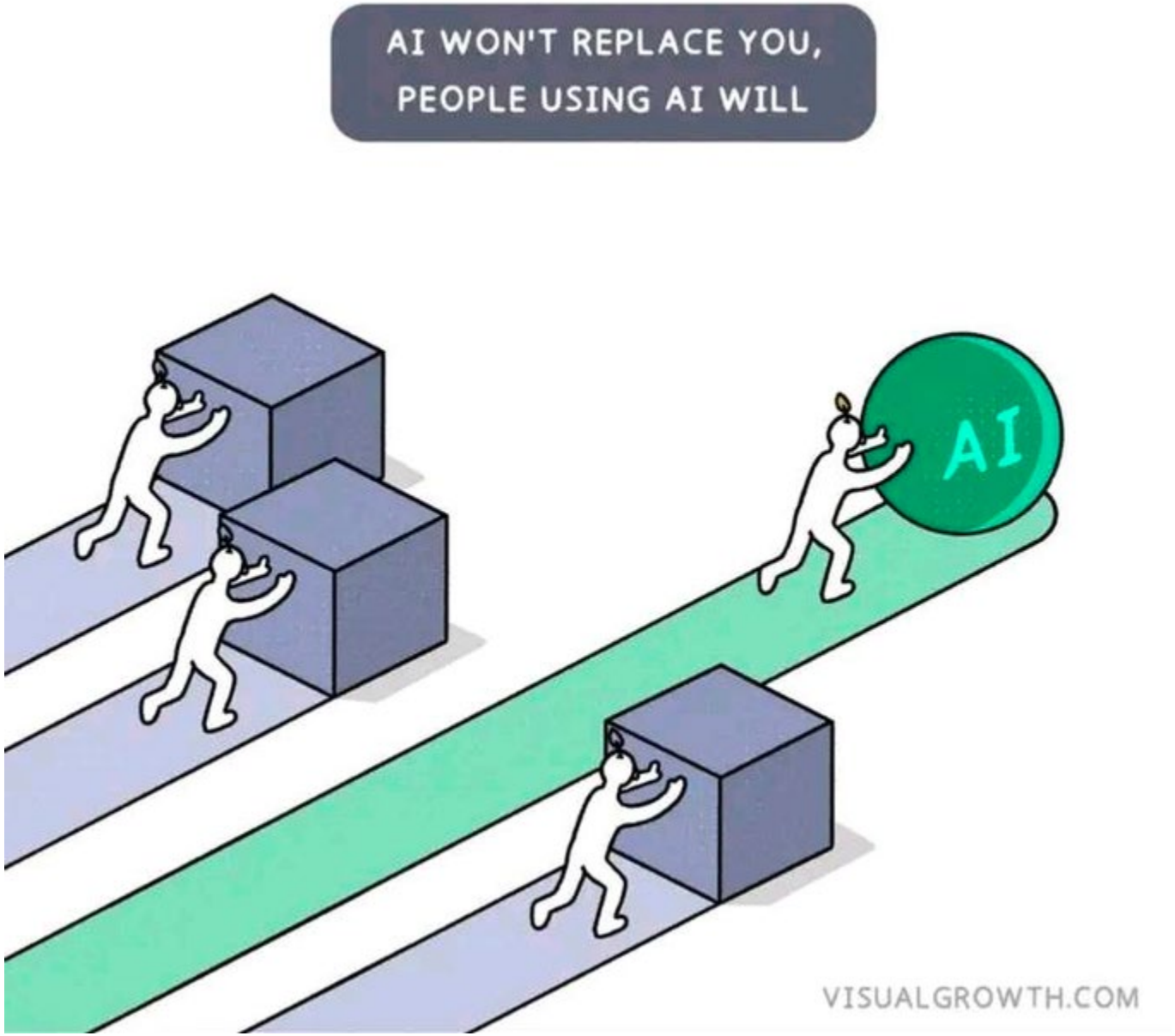
Karl Vandepoele

Molecular Biology & Cytometry Course
08/02/2024

UZ GENT
UNIVERSITEIT GENT

14 / 02 / 1924

Computing Tabulating Recording Company



Rob Thomas, IBM

design qPCR primers and probes to detect FLT3-ITDs

Designing qPCR primers and probes for the detection of FLT3-ITDs (Internal Tandem Duplications in the FMS-like tyrosine kinase 3 gene) requires a good understanding of the gene structure and the specific mutations you are trying to detect. Here's a general approach:

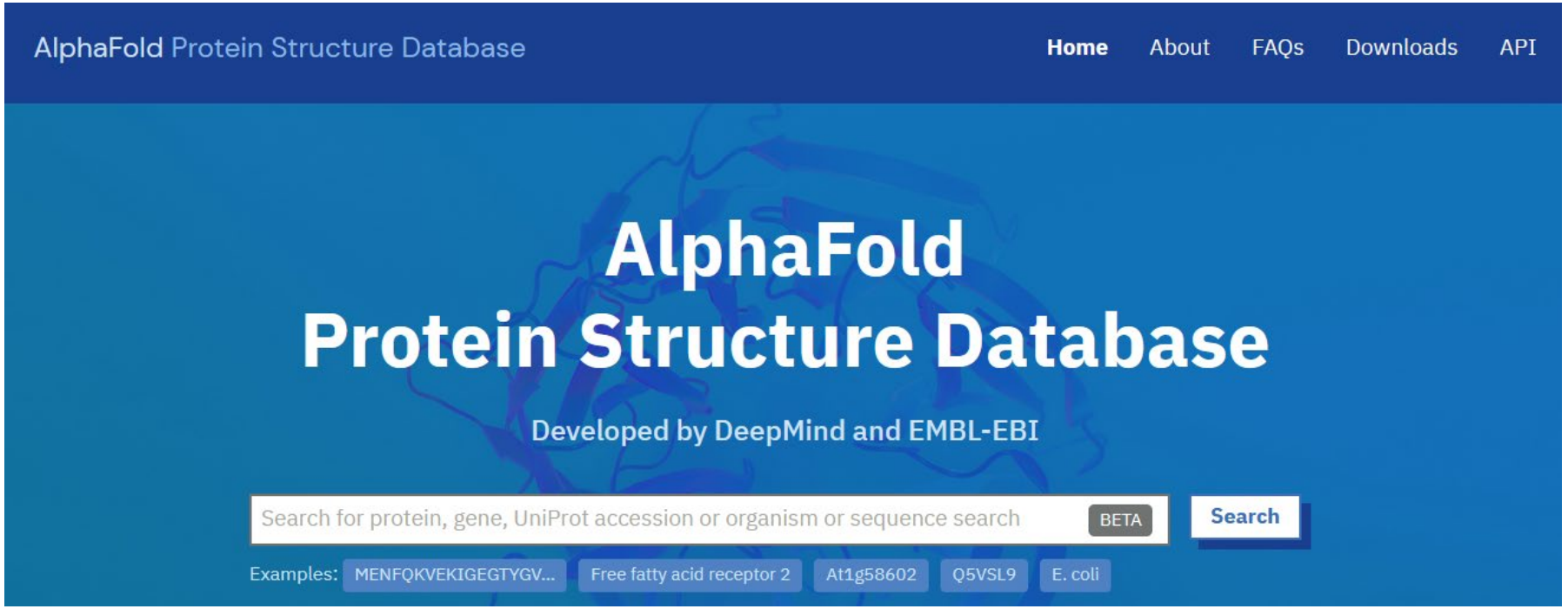
1. **Identify the target sequence**: FLT3-ITDs occur in the juxtamembrane domain-coding sequence of the FLT3 gene, typically between exons 14 and 15. You should obtain the reference sequence for this region.
2. **Design primers and probe**: Use a primer design software to generate primer and probe sequences that are specific to the FLT3 sequence. The primers should flank the region where ITDs commonly occur. The probe should span the junction between the two exons to ensure it will not bind to the duplicated sequence in case of an ITD.
3. **Check for specificity**: Use a tool like BLAST to check the specificity of your primers and probe. They should only bind to the FLT3 gene and not to other sequences in the genome.
4. **Optimize the qPCR conditions**: This includes optimizing the annealing temperature and the primer and probe concentrations.

Please note that this is a general guide and the actual design might need to be adjusted based on the specific requirements of your experiment. Always follow the guidelines provided by the manufacturer of your qPCR reagents and equipment.

Here's an example of what the primer and probe sequences might look like:

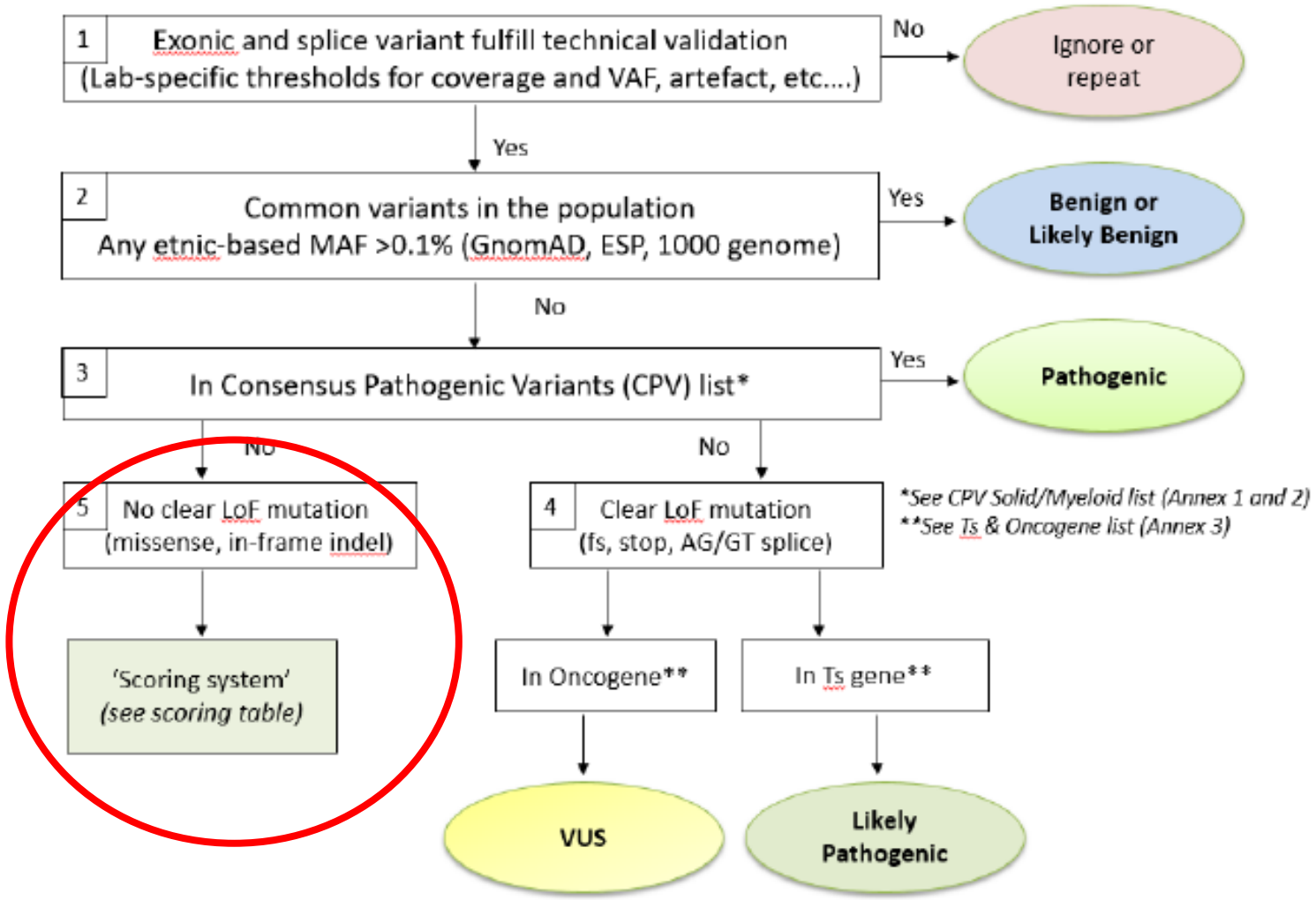
```
Forward Primer: 5'-CTTCTGTTTCAGAAGGAAGC-3'
Reverse Primer: 5'-CTTCTGTTTCAGAAGGAAGC-3'
Probe: 5'-FAM-AGGATTTGGGCTAAAGCAG-BHQ1-3'
```

Remember, these are just hypothetical sequences and should not be used without proper validation. Always consult with a molecular biologist or a similar expert when designing experiments.



- Launched in 2020 with 100 000 unique protein structures (Jumper et al., Nature 2020)
- Human proteome (98.5%) published in 2021 (Tunyasuvunakool et al., Nature 2021) => Method of the year 2021
- Latest database release > 200 million entries

- Future developments: atomic scale modeling of interaction between proteins and nucleic acids



Froyen et al., Cancers 2019

4 million observed missense variants

only 2% classified as benign/pathogenic

(Cheng et al., Science 2023)

RESEARCH ARTICLE SUMMARY

MACHINE LEARNING

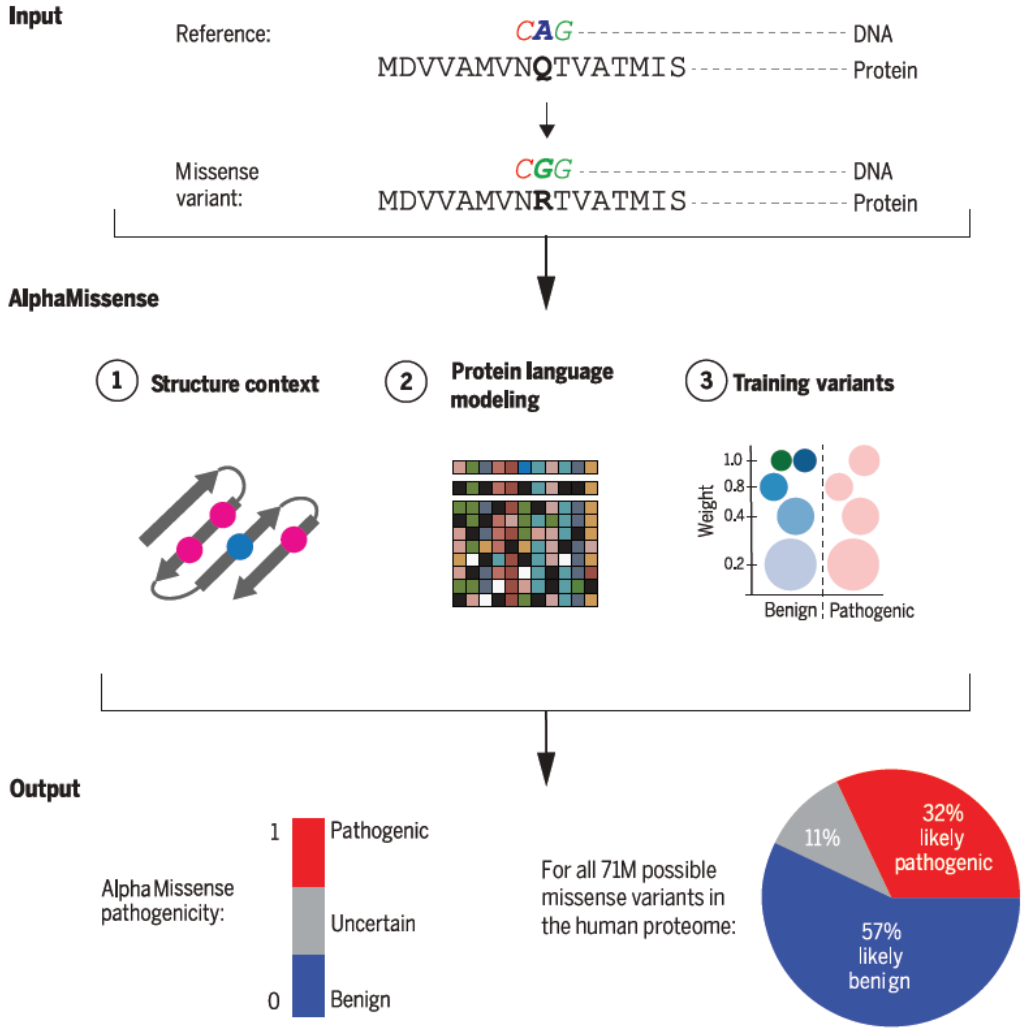
Accurate proteome-wide missense variant effect prediction with AlphaMissense

Jun Cheng*, Guido Novati, Joshua Pan†, Clare Bycroft†, Akvilė Žemgulytė†, Taylor Applebaum†, Alexander Pritzel, Lai Hong Wong, Michal Zielinski, Tobias Sargeant, Rosalia G. Schneider, Andrew W. Senior, John Jumper, Demis Hassabis, Pushmeet Kohli*, Žiga Avsec*

Cheng *et al.*, *Science* **381**, 1303 (2023) 22 September 2023

- AlphaMissense: Combination of structural context and evolutionary conservation to predict pathogenicity

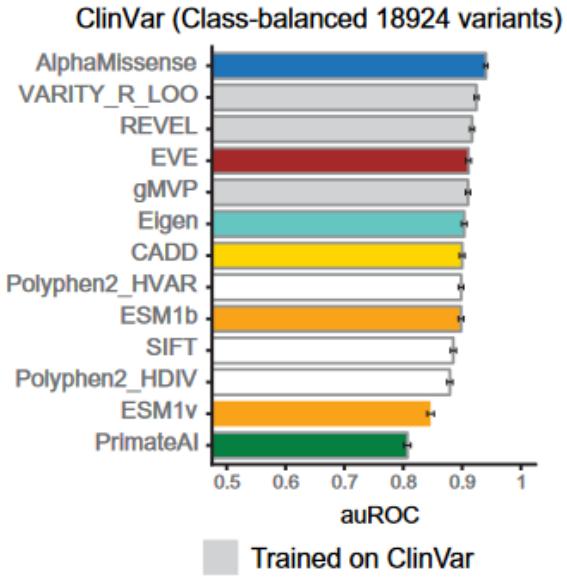
- Not trained on human classification => no human bias
  - Incorporating structural context by using an AF-derived system
  - Unsupervised protein language modeling to learn amino acid distribution depending on sequence context
  - Fine-tuning on weak labels from population frequency data

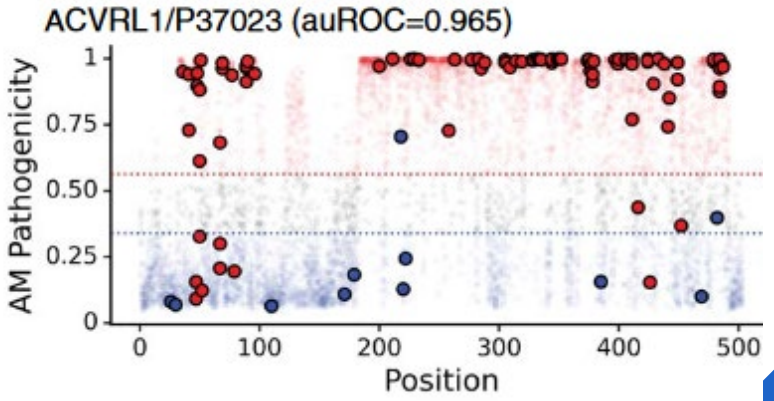


89% classified

AlphaMissense

- Validated on ~19 000 missense variants in ClinVar :
  - auROC: 0,94
- Database of precomputed 71 million missense variants (5 Gb file)
- Only for single nucleotide variants (SNV)
- Indels: 22% of disease-associated mutations (Human Gene Mutation Database)





dots: ClinVar entries

Real life evaluation of AlphaMissense predictions in hematological malignancies

Kaddour Chabane[1], Carole Charlot[1], Dan Gugenheim, Thomas Simonet[2], David Armisen[3], Pierre-Julien Viailly[4], Guillaume Codet de Boisse[5], Sarah Huet[1,3], Sandrine Hayette[1], Vincent Alcazer[3,6] and Pierre Sujobert[1,3]✉

Leukemia; https://doi.org/10.1038/s41375-023-02116-3

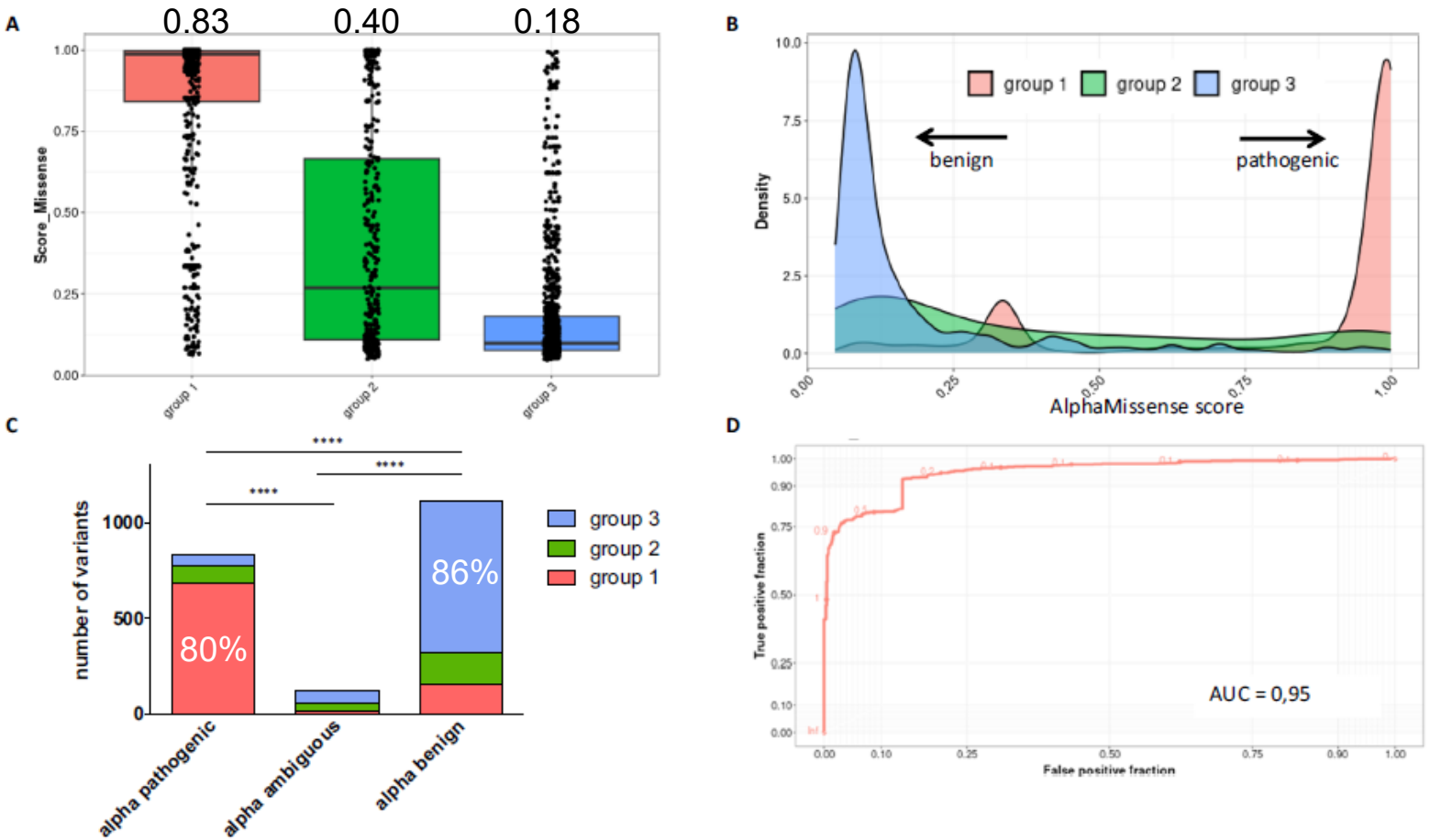
- Dataset of 2222 missense variants from patients with hematological malignancies
  - 2073 predictions by AM (93.3%)
  - AM uses UniProtKB annotation

- Group 1: Variants with clinical significance (n = 853)
- Group 2: Variants of unknown clinical significance (n = 295)
- Group 3: Variants deemed benign of likely benign (n = 925)

Real life evaluation of AlphaMissense predictions in hematological malignancies

Kaddour Chabane[1], Carole Charlot[1], Dan Gugenheim, Thomas Simonet[2], David Armisen[3], Pierre-Julien Viailly[4], Guillaume Codet de Boisse[5], Sarah Huet[1,3], Sandrine Hayette[1], Vincent Alcazer[3,6] and Pierre Sujobert[1,3]

Leukemia; https://doi.org/10.1038/s41375-023-02116-3



Real life evaluation of AlphaMissense predictions in hematological malignancies

Kaddour Chabane[1], Carole Charlot[1], Dan Gugenheim, Thomas Simonet[2], David Armisen[3], Pierre-Julien Viailly[4], Guillaume Codet de Boisse[5], Sarah Huet[1,3], Sandrine Hayette[1], Vincent Alcazer[3,6] and Pierre Sujobert[1,3]

- 170 Group 1 variants classified as benign by AM
  - 122 with functional evidence for pathogenicity (including JAK2 V617F (93x), MPL, FLT3, UBA1, …)
    - AM false negatives
  - BRAF A762V reclassified as likely benign
- 155 Group 3 variants classified as pathogenic by AM
  - 32 variants identified as polymorphisms (mAF > 0.1% but < 1%): mean allele frequency ~50%
    - AM false positives
  - 11 variants documented in cBioPortal
    - Current pipeline false negatives

- Available as online portal: https://alphamissense.calym.eu/index.php

Franklin by genoox

JAK2:c.1849G>T

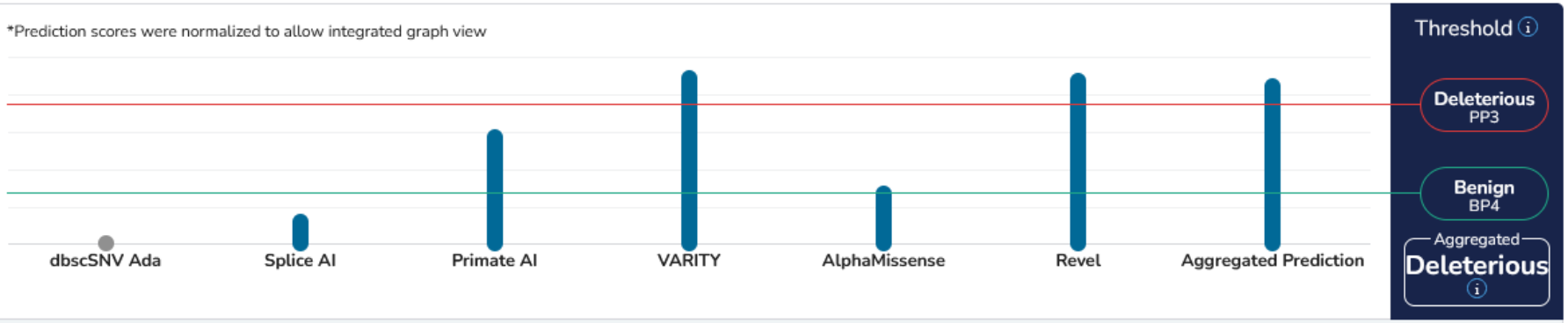
chr9-5073770 G>T | p.Val617Phe | NM_004972.4 | dbSNP, rs77375493 | UCSC, TraP, gnomAD

Somatic Clinical Evidence | Computed Classification | Clinical Trials | Variant Assessment | Gene Assessment | Publications

- Region Viewer
- Franklin Community Frequency
- Somatic Frequency
- Internal Frequency
- Clinical evidence
- **Predictions**
- Population Frequencies
- Transcripts

Predictions

*Prediction scores were normalized to allow integrated graph view



Aggregated	
Aggregated Prediction	Deleterious (0.87)
Functional Coding	
Revel	Deleterious (Moderate) (0.88)
AlphaMissense	Uncertain (0.334)
Eve	(N/A)
Varity	Deleterious (0.88)
MUT Assesor	Med (2.09)
SIFT	Uncertain (0.002)
Polyphen2	Deleterious (Supporting) (0.98)

nature reviews microbiology

https://doi.org/10.1038/s41579-023-00984-1

Review article

Check for updates

Machine learning for microbiologists

Francesco Asnicar[1,5], Andrew Maltez Thomas[1,5], Andrea Passerini[2], Levi Waldron[1,3] & Nicola Segata[1,4]

Abstract

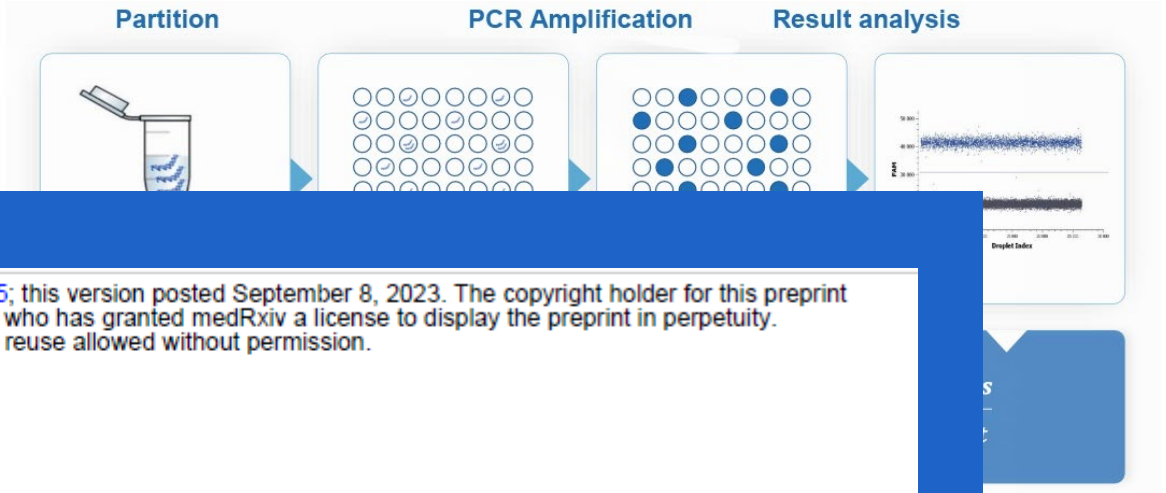
Machine learning is increasingly important in microbiology where it is used for tasks such as predicting antibiotic resistance and associating human microbiome features with complex host diseases. The applications in microbiology are quickly expanding and the machine learning tools frequently used in basic and clinical research range from classification and regression to clustering and dimensionality reduction. In this Review, we examine the main machine learning concepts, tasks and applications that are relevant for experimental and clinical microbiologists. We provide the minimal toolbox for a microbiologist to be able to understand, interpret and use machine learning in their experimental and translational activities.

Sections

Digital PCR meets High Resolution Melting (HRM)

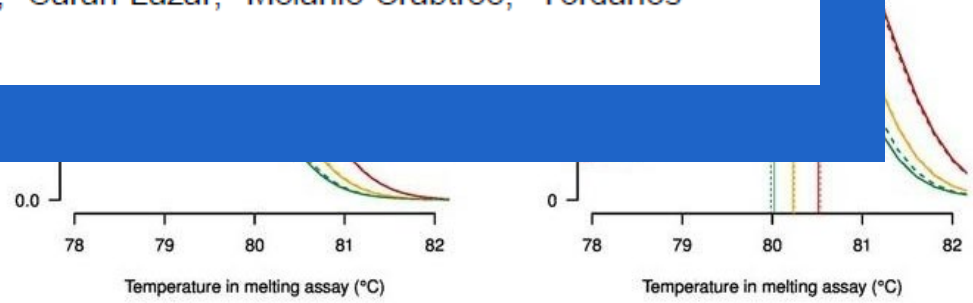
Use of AI in microbiology

- Digital PCR:
  - P
  - C
- High
  - P
  - M



Universal digital high resolution melt analysis for the diagnosis of bacteremia

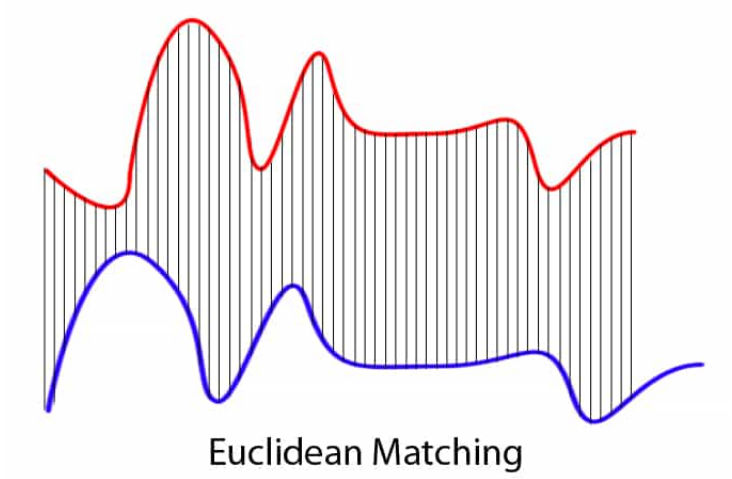
April Aralar,[a] Tyler Goshia,[a] Nanda Ramchandar,[b,c] Shelley M. Lawrence,[d] Aparajita Karmakar,[e] Ankit Sharma,[e] Mridu Sinha,[e] David T. Pride,[f] Peiting Kuo,[f] Khrissa Lecrone,[f] Megan Chiu,[f] Karen Mestan,[g] Eniko Sajti,[g] Michelle Vanderpool,[h] Sarah Lazar,[g] Melanie Crabtree,[g] Yordanos Tesfai,[g] Stephanie I.Fraley[a*]#

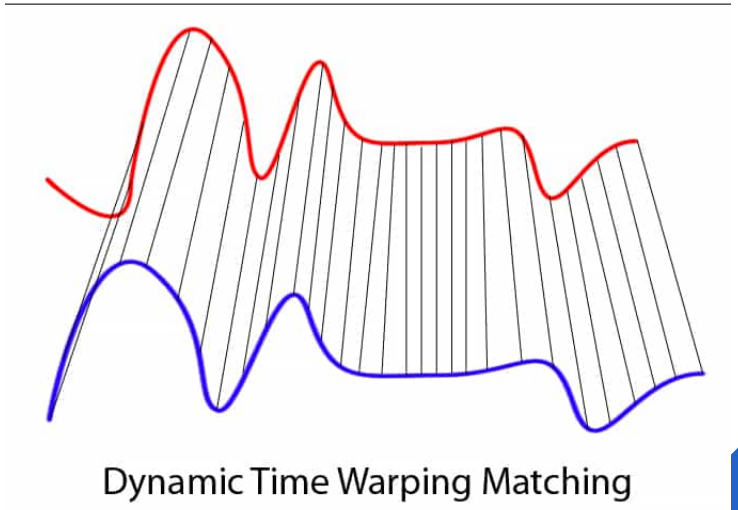


Digital PCR meets High Resolution Melting (HRM)

Use of AI in microbiology (Aralar et al., medRxiv 2023; Goshia et al., bioXriv 2023)

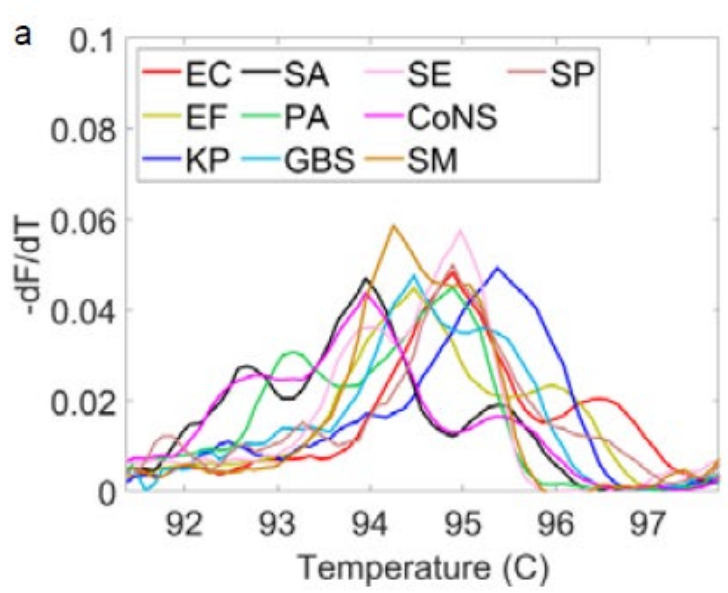
- Fast and accurate diagnosis of bloodstream infection
  - Blood culture: gold standard (~15h, but can take multiple days)
- Instrument: self-developed (see previous publications from this group)
- 16S rDNA => database of melting curves for 11 bacterial species
  - ML algorithm combining dynamic time warping and Euclidean distance
  - Optimized image processing, melt curve preprocessing and machine learning pipeline
    - K-Means clustering to extract key clusters
    - k-Nearest Neighbour classifier to classify a test curve

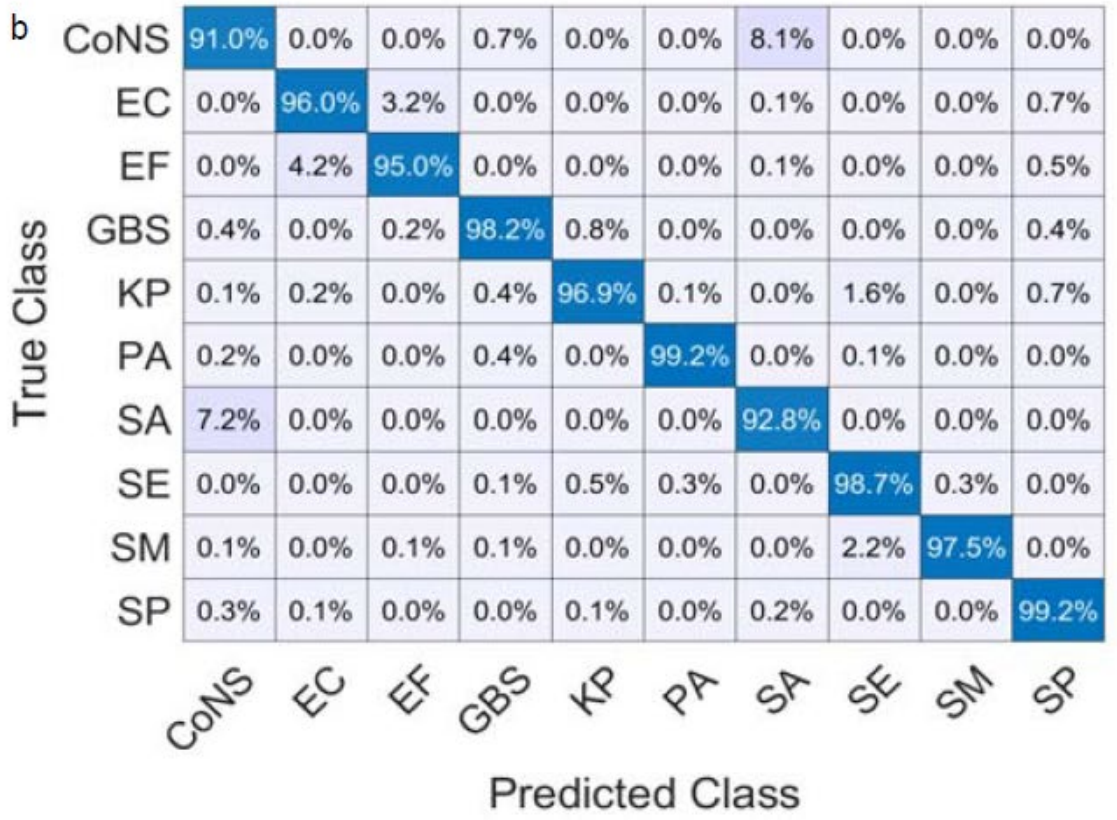




Digital PCR meets High Resolution Melting (HRM)

Training the classifier

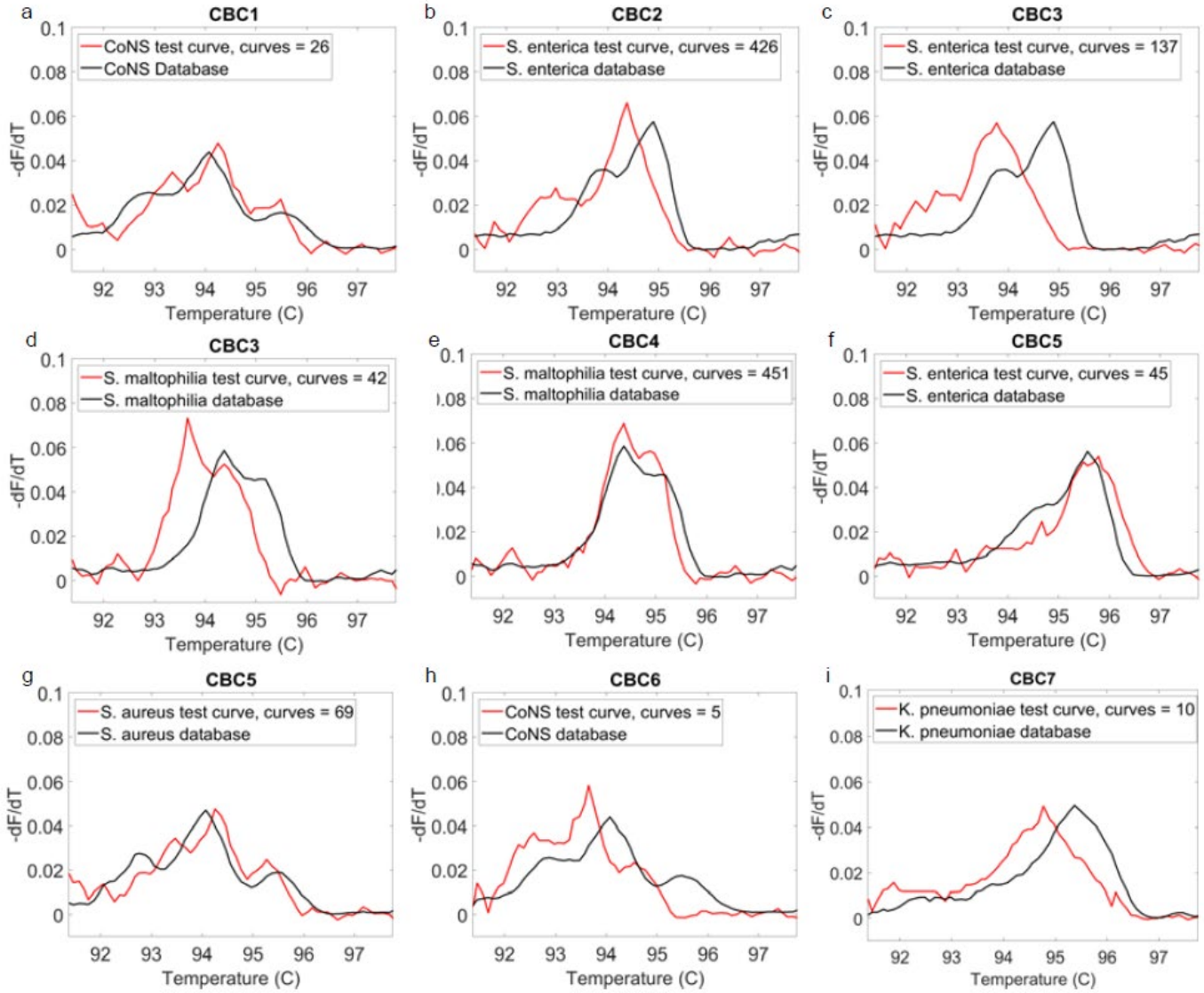




97% classification accuracy
> 146,000 training curves

Digital PCR meets High Resolution Melting (HRM)

Using real patient samples



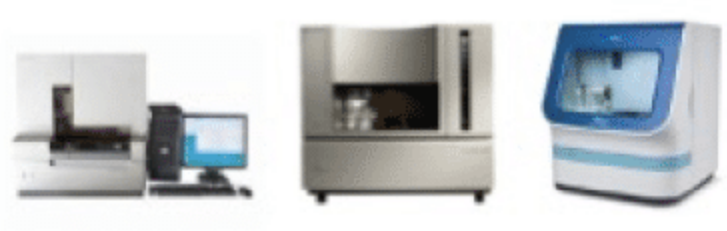
U-dHRM

Exact quantification

100% correct species identification

Short TAT (6h)

First generation



Sanger sequencing
Maxam and Gilbert
Sanger chain termination

Infer nucleotide identity using dNTPs, then visualize with electrophoresis

500–1,000 bp fragments

Second generation
(next generation sequencing)

454, Solexa,
Ion Torrent,
Illumina

High throughput from the parallelization of sequencing reactions

~50–500 bp fragments

Third generation

PacBio
Oxford Nanopore

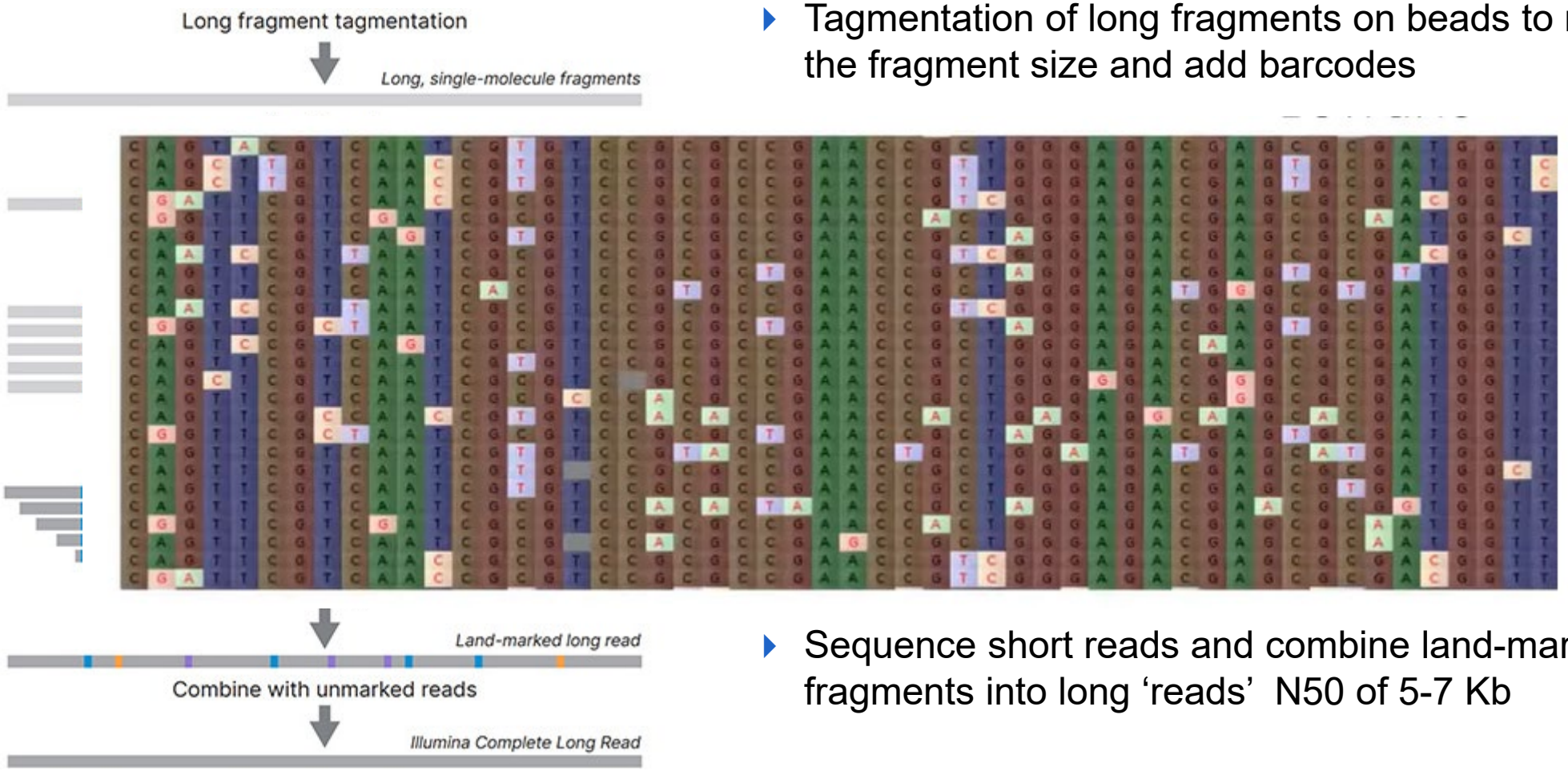
Sequence native DNA in real time with single-molecule resolution

Tens of kb fragments, on average

Short-read sequencing

Long-read sequencing

Illumina long read sequencing



- Tagmentation of long fragments on beads to normalize the fragment size and add barcodes

·range
s (Longas

ary wtih

- Sequence short reads and combine land-marked fragments into long ‘reads’ N50 of 5-7 Kb

Applications of Illumina long read sequencing

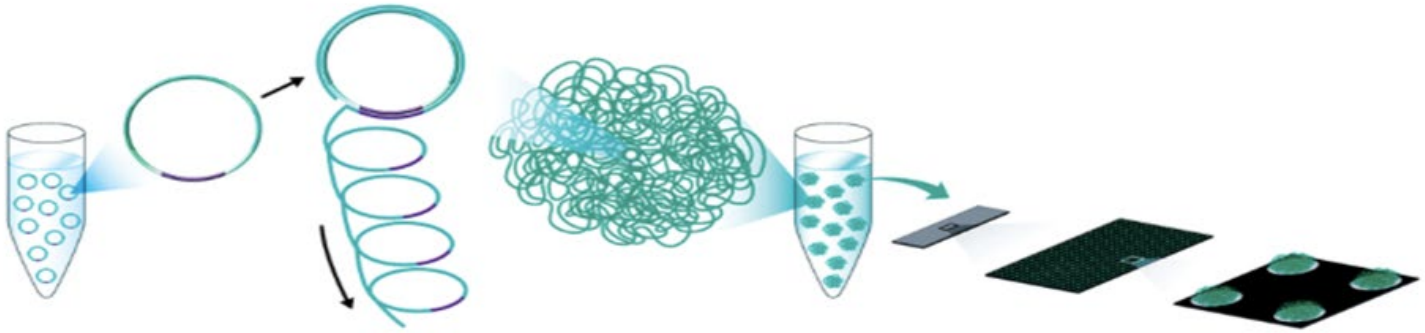
- Illumina Complete Long Reads can resolve highly polymorphic regions like the *HLA-A* gene. Uniform coverage in the *HLA* region enables accurate phasing of *HLA* alleles.
- A heterozygous 180 bp deletion in the *SEMG1* gene is clearly detected by both Illumina Complete Long Reads and on-market long reads.

Detection of clinical pathogens

- Fast and accurate detection of pathogens is crucial (remember SARS-CoV-2)
  - Protein-based diagnostics: easy to scale up, simple to use BUT rely on high quality antibody => take time
  - Nucleic acid-based diagnostics: easier to develop, higher sensitivity and flexibility

- LAMP: loop-mediated isothermal amplification for simple implementation and scalability
  - 4-6 oligonucleotides targeting a region of interest + strand-displacing polymerase
  - No need for thermocycler
  - Detection by gel electrophoresis, fluorescent, turbidity or colorimetric methods => cost increases depending on method used or risk of false positives

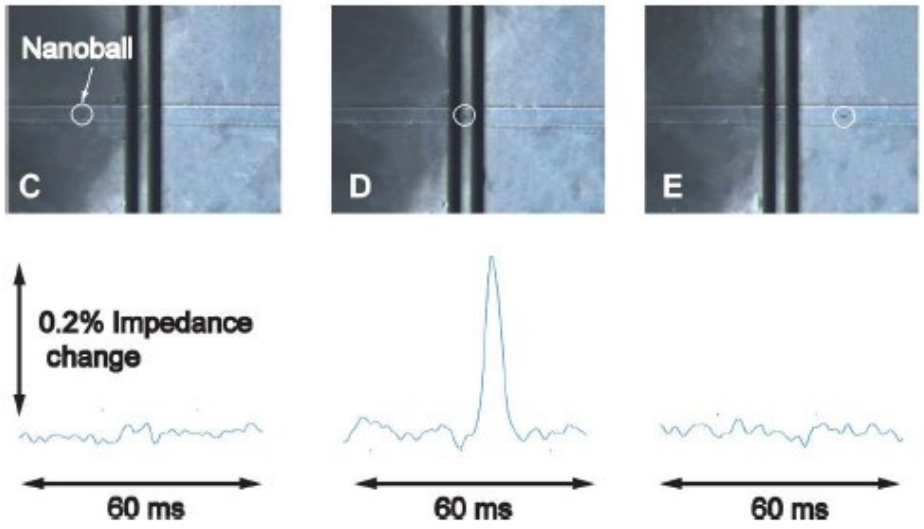
- DNA nanoballs: uses rolling-circle amplification, used for sequencing

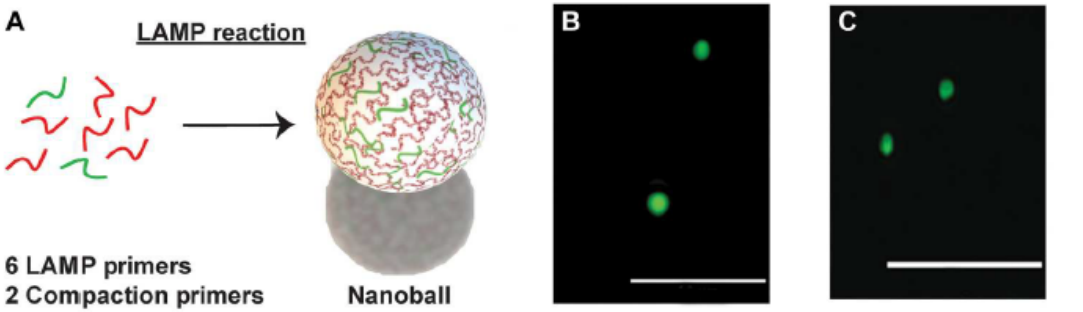


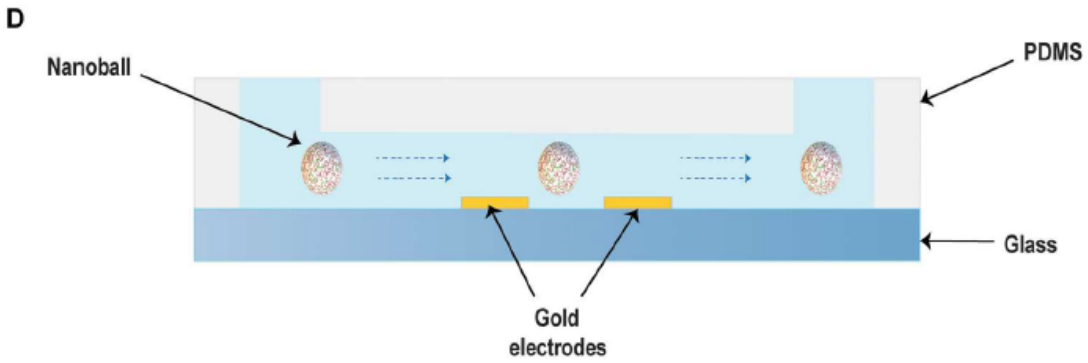
Detection of clinical pathogens

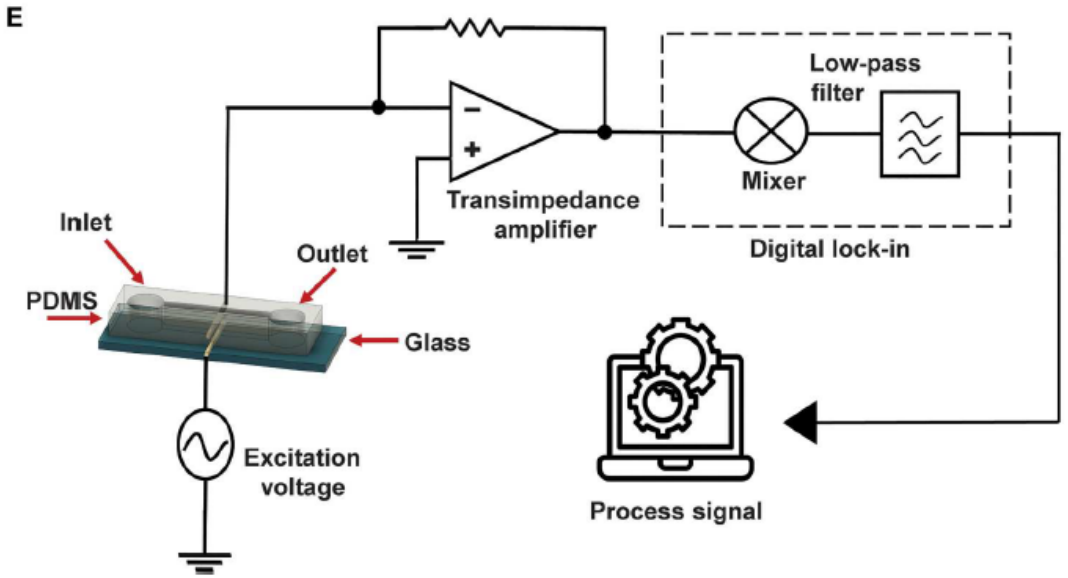
- How to make DNA nanoballs from shorter LAMP fragments?
  - (Use microbeads as external nucleation agents)
  - Use oligonucleotides complementary to common region in amplicons to “staple” them together
    - Oligo’s already present in LAMP reaction

- Microfluidic system to detect presence of DNA nanoballs

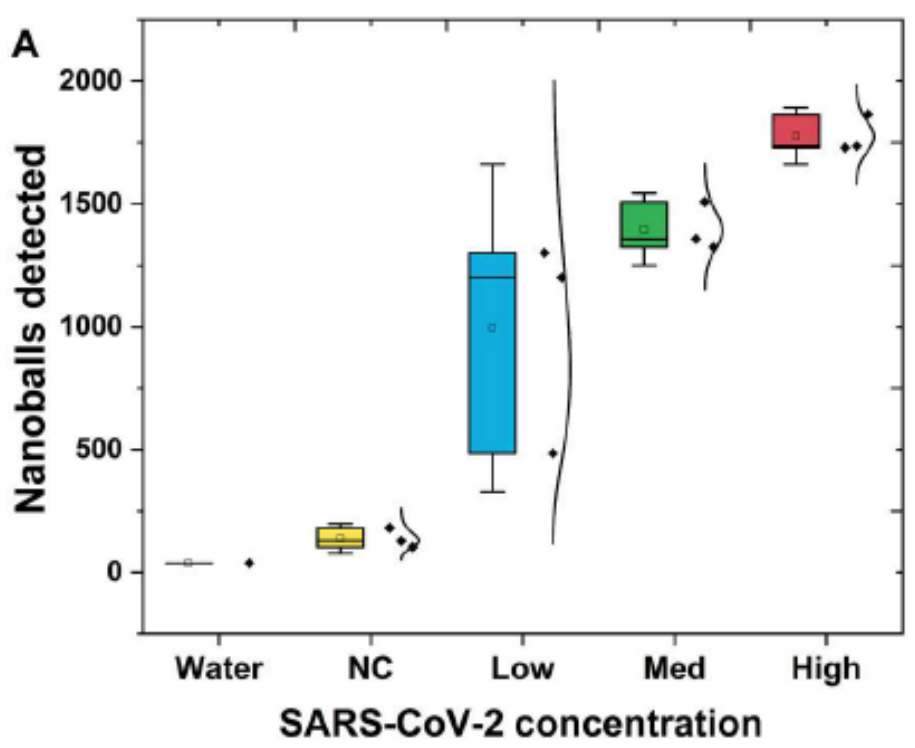


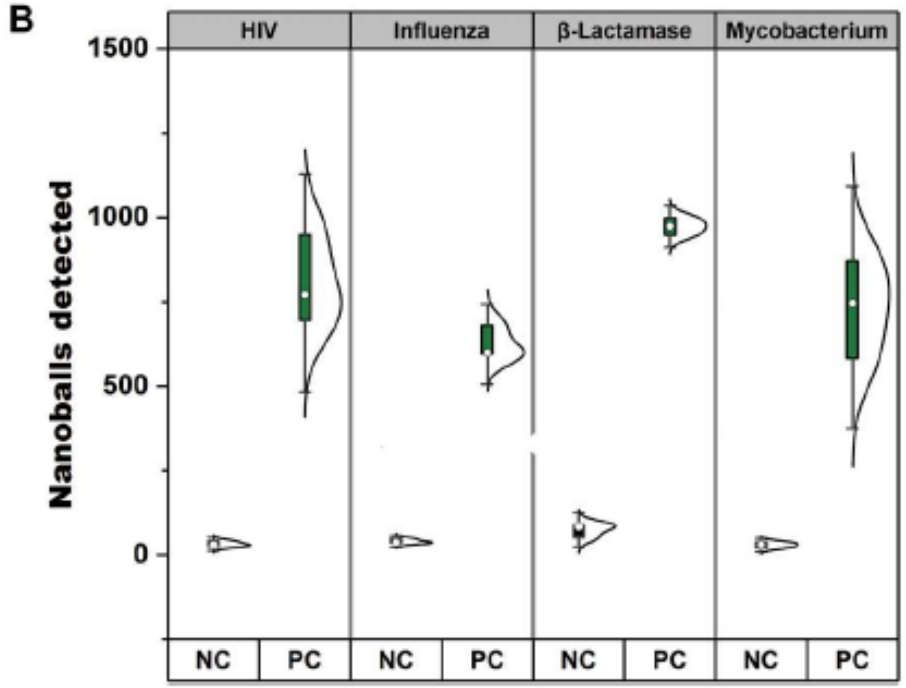






Detection of clinical pathogens





SCIENCE ADVANCES | RESEARCH ARTICLE

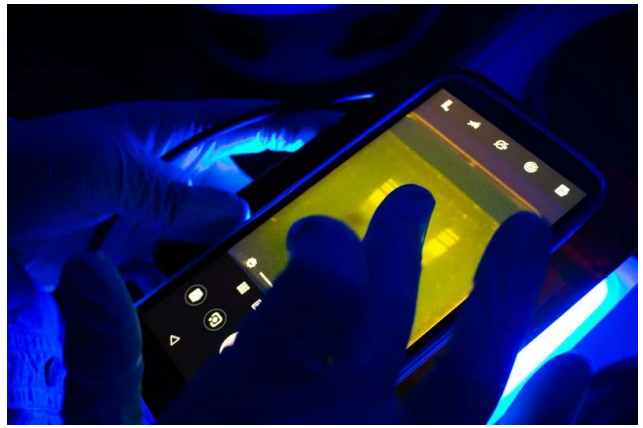
APPLIED SCIENCES AND ENGINEERING

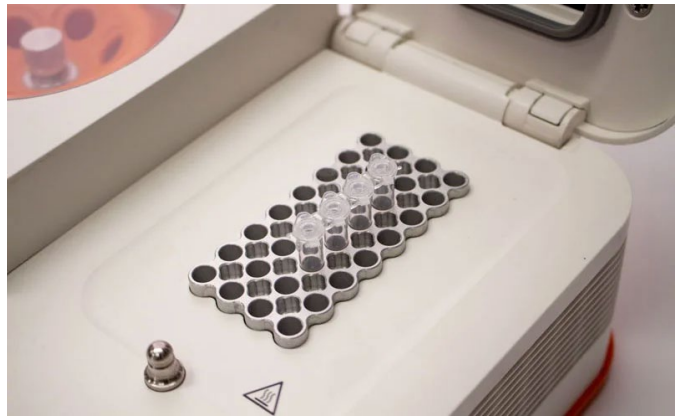
Tayyab *et al.*, *Sci. Adv.* **9**, eadi4997 (2023) 6 September 2023

Digital assay for rapid electronic quantification of clinical pathogens using DNA nanoballs

Muhammad Tayyab[1]†, Donal Barrett[2]†, Gijs van Riel[2], Shujing Liu[2,3], Björn Reinius[4], Curt Scharfe[5], Peter Griffin[6], Lars M. Steinmetz[6,7]*, Mehdi Javanmard[1]*, Vicent Pelechano[2]*

4-well heating block
Power delivered by USB-C port
Store 16 protocols in memory
80 g
250 $





1459 € / 1799 € for basic/pro version
Bento Lab Devices

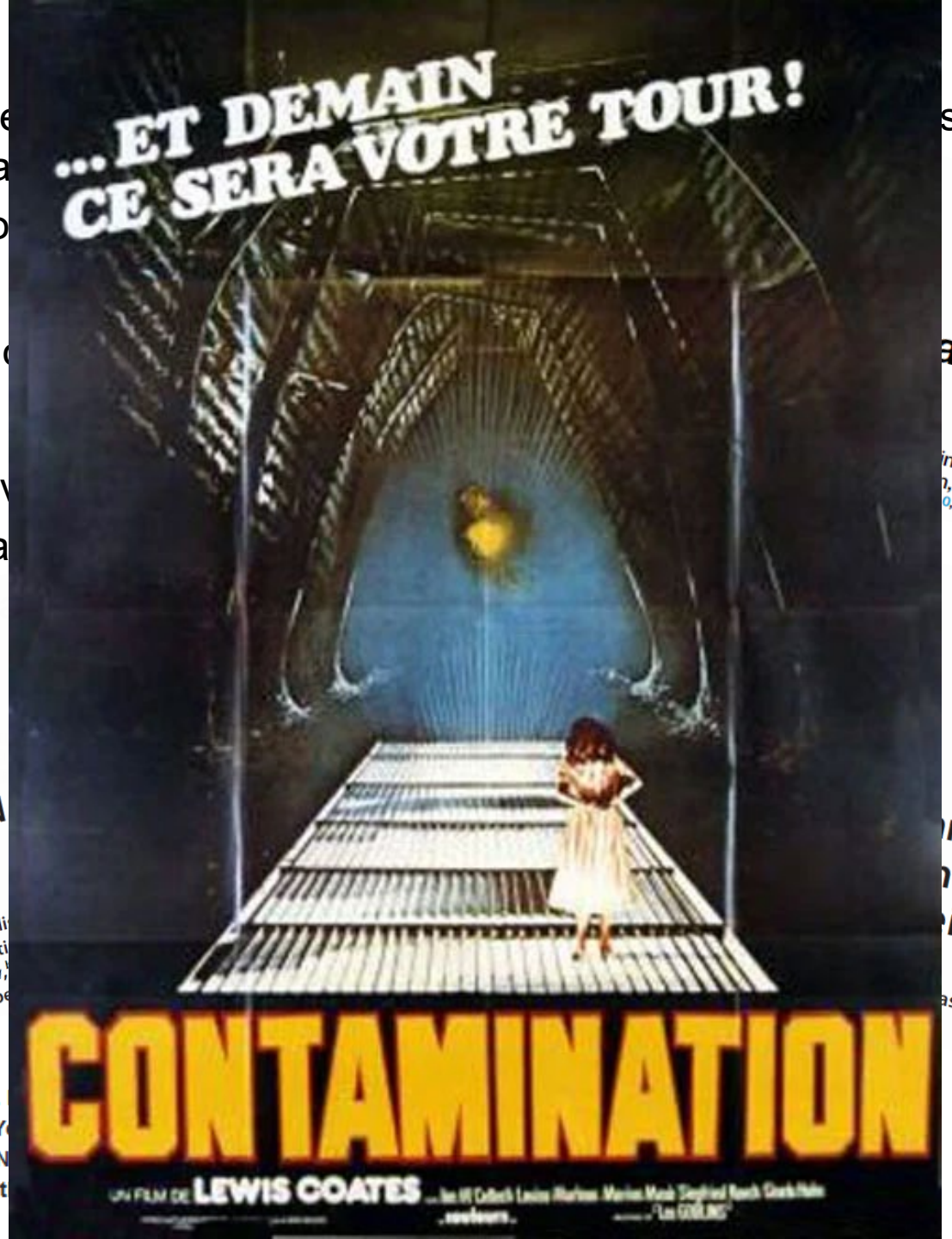
June 2020

- Four colleagues in Seattle, WA (USA) test positive for SARS-CoV-2 at enrollment in a study to determine biomarkers predictive of SARS-CoV-2 symptoms
- High Ct-values N1 38.5 (range 32.5 – 42.8) and N2 39.3 (range 33.2 – 44.1)
- No known exposures to SARS-CoV-2, 10 days of isolation and no symptoms appearing
- No SARS-CoV-2 antibodies detected in plasma at enrollment or after 4 weeks of follow-up
- Persistently testing positive for multiple weeks

- Unable to detect regions of SARS-CoV-2 outside the nucleocapsid ?

- People worked in a biomedical lab and had worked with a plasmid vector in the month prior to enrollment
  - PCR without RT step => positive
  - PCR for AMP or KAN/NEO resistance genes => POS
  - PCR for CMV promoter with codon-optimized sequence => POS in 2/5
    - Sequencing confirmed identity as plasmid-derived

- Samples at enrollment we … s completed before they enrolled in the study & sa…
  - Low risk of sample co…
- One household member …
- Samples remained positiv…
  - Plasmid present in na…
  - *E. coli* with pla…



AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Spectrum

Laboratory-Generated DN… Diagnostic Test Results

Lindsey R. Robinson-McCarthy,[a] Alexander J. Mijali… Robert A. Rasmussen,[b] Raphael Ferreira,[a,b] Jeanti… Erkin Kuru,[a,b] Adama M. Sesay,[b] Joshua Rainbow,… Devora Najjar,[b,d,i] Peng Yin,[b,c] Donald E. Ingbe…

Ingrid A. … Winnie Y… Jiang[a], N… Noah Sat…

CellPress OPEN ACCESS

…atories masquerade

…ine M. Klapperich,[5] Lena Landaverde,[5] …n,[2,7] Judy T. Platt,[8] Kayla Kuhfeldt,[8] …0,11,12 and Michael Springer[4,*]

LETTER TO THE EDITO…

Check for updates

…mination with Noninfectious …n False-Positive Reverse …illance Testing

…ason W. Botten,[c,d,e] Jessica W. Crothers,[d,f]

Questions?