Exploiting the potential of high-throughput sequencing

Jean-Baptiste Veyrieras

BIOMINING LABS

BIOMÉRIEUX

Data & Knowledge Lab.

Outline

- Are you ready to rescale your mind ?
- HT sequencing technologies
- Current bottlenecks for clinical/industrial applications
- A versatile technology
- Conclusion

New Era, New Paradigm

- **Advent of affordable high-throughput (HT) life-science technologies in biology have rescaled the research landscape**

- **Technological Revolution**
  - ✓ **HT Genomics (HT sequencing)**
    - Pan/Quantitative Genomics
  - ✓ **HT Epigenomics (micro-arrays, HT sequencing)**
    - Pan/Quantitative EpiGenomics
      - DNA Methylation, Chromatin structure
  - ✓ **HT Transcriptomics (micro-arrays, HT sequencing)**
    - Pan/Quantitative Transcriptomics
      - Coding & non-coding RNA
  - ✓ **HT Proteomics/Metabolomics (mass-spectrometry)**
    - Pan/Quantitative Proteomics/Metabolomics

New Era, New Paradigm

- **What impact these technologies can have on medical practices ?**

- **Medical Revolution / Personalized Medecine (4P)**
  - ✓ **Predictive**
    - early diagnostic
  - ✓ **Personalized**
    - diagnostic, treatment | patient
  - ✓ **Preemptive**
    - preempt disease before occurrence
  - ✓ **Participatory**
    - individual + institutions (network)

New Era, New Paradigm

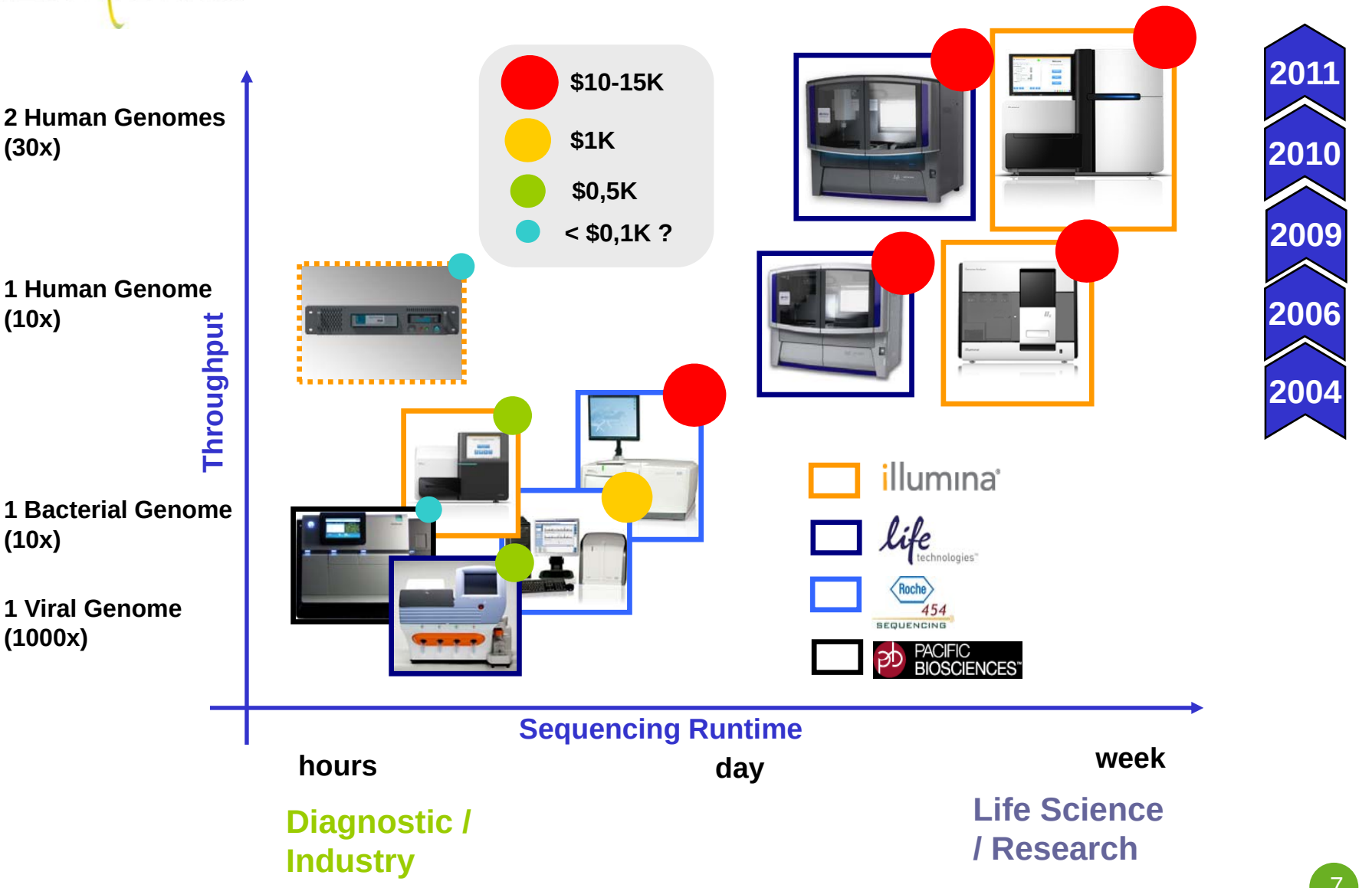
- **Web 2.0 leverage data & knowledge sharing in science**
  - ✓ **Network/Collaborative Era**
    - BioWikis/BioTorrent/BioSharing
  - ✓ **3.0 revolution is coming**
    - Semantic Web (Linked Data / Ontologies)

- **DNA is a digital information**
  - ✓ **Genotypic vs phenotypic data (e.g. mass-spectrometry)**
    - Dirac (Digital) vs Gaussian (Analogic)
  - ✓ **DNA can be shared/reused easily**
    - Incremental learning
      - No worry about heterogeneity between studies

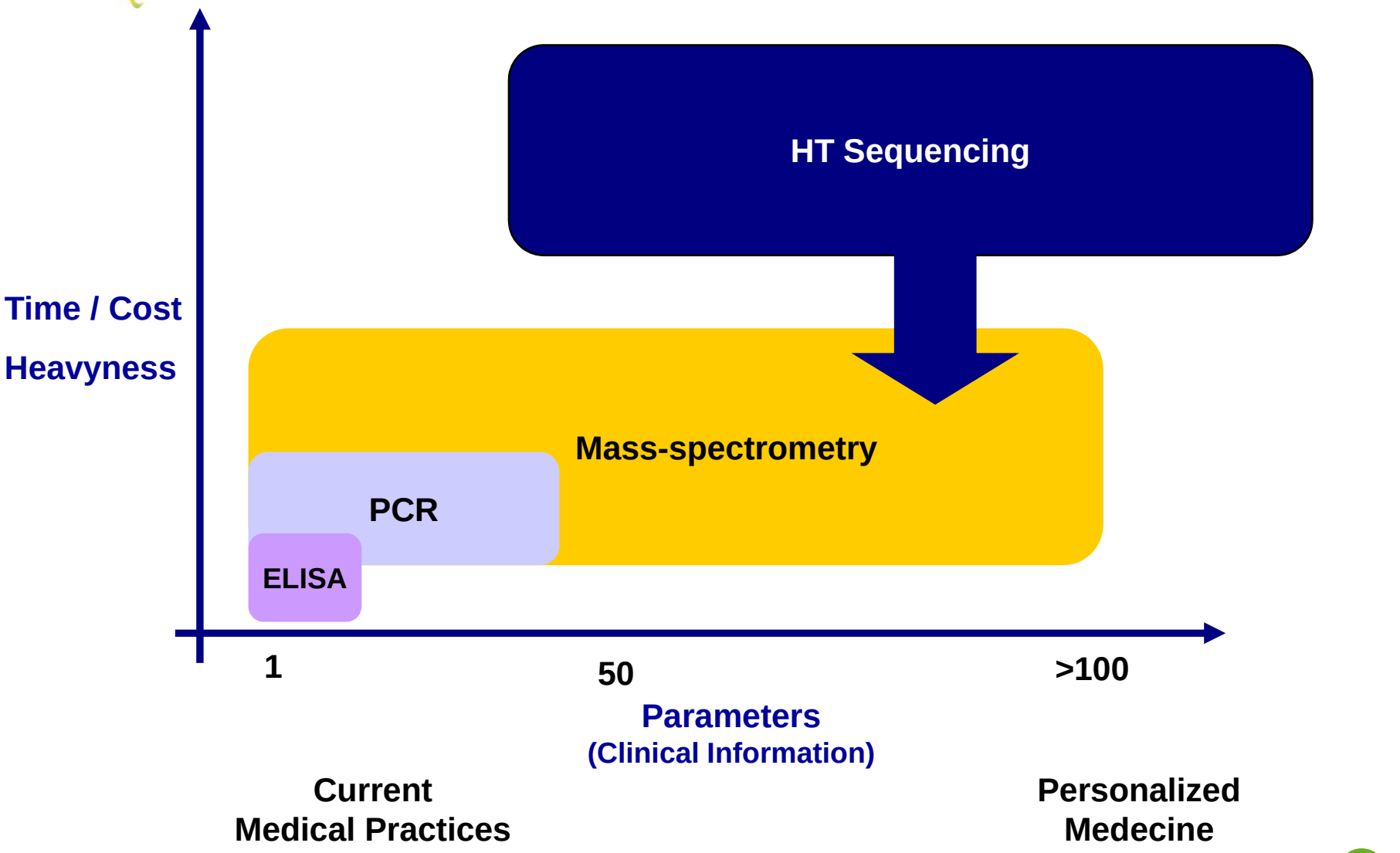
Outline

- Are you ready to rescale your mind ?
- **HT sequencing technologies**
- Current bottlenecks for clinical/industrial applications
- A versatile technology
- Conclusion

A cartography of HT sequencing instruments



HT Sequencing & IVD technologies



Outline

- Are you ready to rescale your mind ?
- HT sequencing technologies
- **Current bottlenecks for clinical/industrial applications**
- A versatile technology
- Conclusion

Pre and post analytics bottlenecks

Pre-analytics → Analyzer → Post-analytics

Clinical / Industrial Samples

Clinical / Industrial Actionable Information

Sample / Target Preparation

From raw sample to ready-to-sequence molecules

Critical items:

Robustness/Faithfulness/Easiness/ Cost/Time

IT System

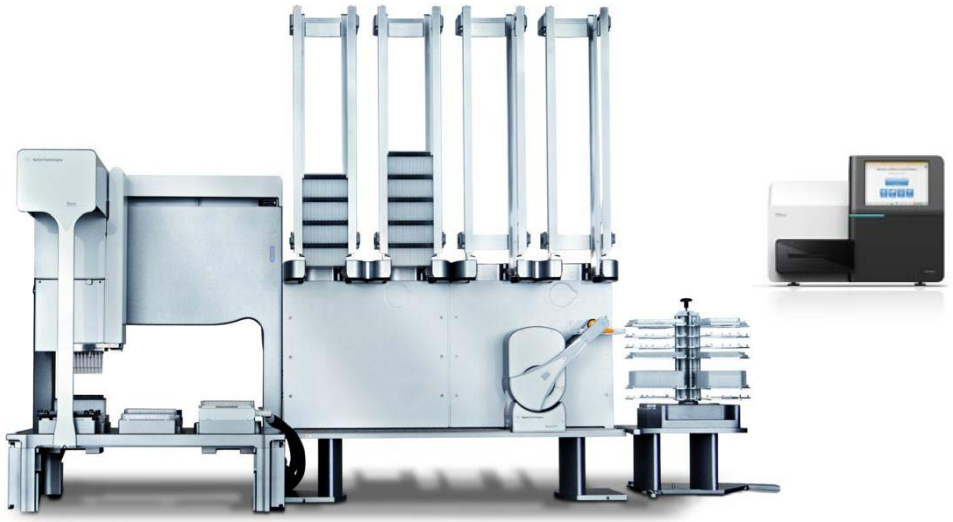
From reads to customer relevant information

Critical items:

CPUs/Time/ Data and Knowledge/Software

Pre and post analytics bottlenecks

Pre-analytics → **Analyzer** → **Post-analytics**



Agilent Technologies

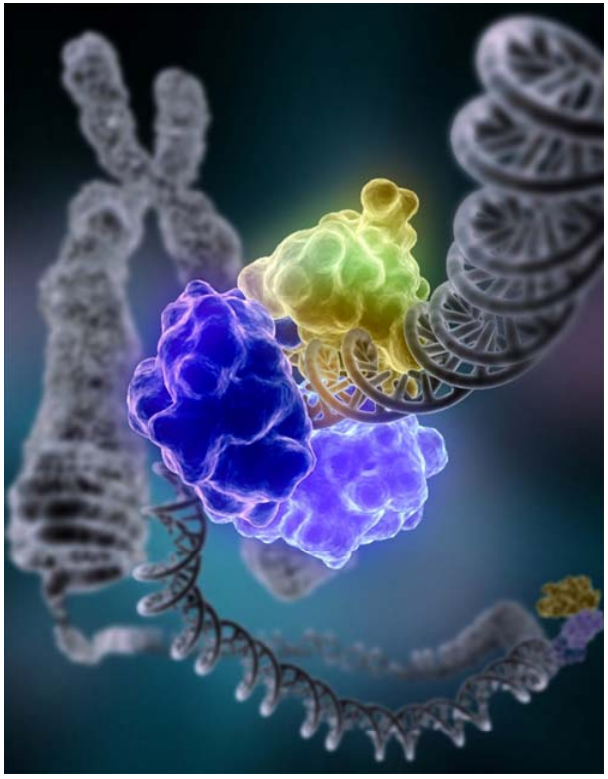
Outline

- Are you ready to rescale your mind ?
- HT sequencing technologies
- Current bottlenecks for clinical/industrial applications
- **A versatile technology**
- Conclusion

Only HT sequencing can do this !

- **An amazing versatility**

  - ✓ **DNA sequencing**
  - ✓ **Protein-DNA binding**
    - DNAse footprinting / ChIP-Seq
  - ✓ **Epigenomics**
    - Methylation (PacBio)
    - Chromatin structure (ChIP-Seq)
  - ✓ **Transcriptomics**
    - RNA-sequencing
  - ✓ **Proteomics !**
    - Aptamer-sequencing



HT Sequencing and Clinical/Industrial end-points

- **Clinical end-point**
  - ✓ **pathogen, disease, treatment response, etc…**
  - ✓ **CONTENT**
    - Improvement of predictive models
      - HT screening of novel predictive features
    - Improve/Refine biological knowledge
      - Transfer learning
  - ✓ **SYSTEM**
    - Affordable cost per sample while maximizing accuracy & personalizing prediction (multiparametrics)
- **Industrial end-point**
  - ✓ **Pharma, Cosmetic, Food, Agronomy, Ecology**
    - Monitoring of product/environment quality
  - ✓ **SYSTEM**
    - New technologies to perform fine-grained HT QC

Example of applications

➢ Infectious Diseases

✓ Inject knowledge to improve/develop

- IVD tests based on existing technologies
- Antimicrobials

✓ Real-time epidemiology

✓ Metagenomics

- A discovery platform for biomarkers
- A present quality-control platform
- A future diagnostic platform ?

Fast IVD development for emerging pathogens

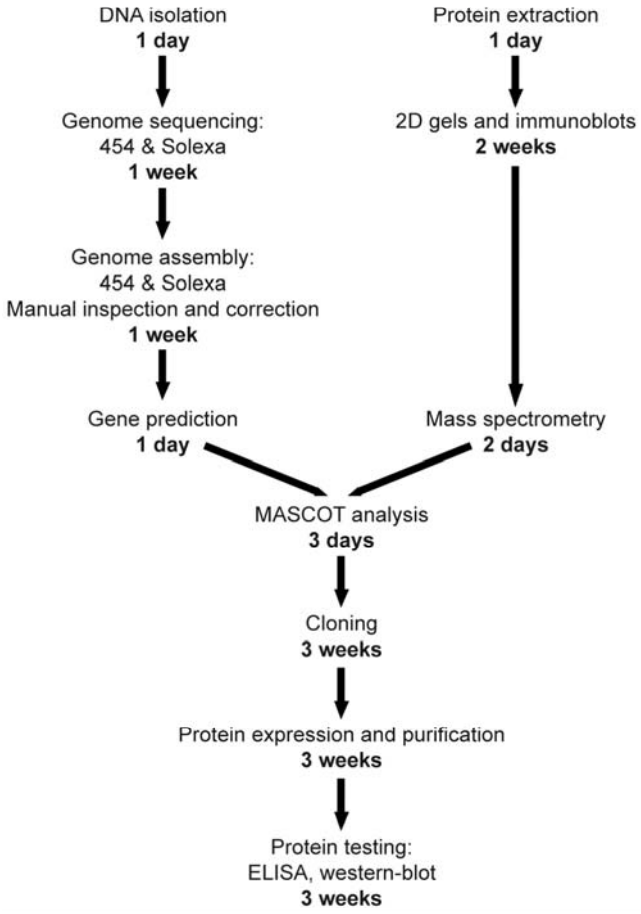
OPEN ACCESS Freely available online

PLoS one

High Throughput Sequencing and Proteomics to Identify Immunogenic Proteins of a New Pathogen: The Dirty Genome Approach

Gilbert Greub[1☯*], Carole Kebbi-Beghdadi[1☯], Claire Bertelli[1☯], François Collyn[1], Beat M. Riederer[2,3], Camille Yersin[1], Antony Croxatto[1], Didier Raoult[4]

1 Center for Research on Intracellular Bacteria (CRIB), Institute of Microbiology, University Hospital Center, University of Lausanne, Lausanne, Switzerland, 2 Department of Cellular Biology and Morphology, University of Lausanne, Lausanne, Switzerland, 3 Proteomics Unit, Department of Psychiatric Neurosciences, Cery, Prilly-Lausanne, Switzerland, 4 Unité des Rickettsies, Faculté de Médecine, Université de la Méditerranée, Marseille, France



- **Isolate & sequence new pathogen**
- **Search for specific immunogenic proteins**
  - ✓ **Comparative proteogenomics**
- **Validation of candidate protein markers**
- **Transfer to routine IVD platform**

Example of applications

➢ Infectious Diseases

✓ Inject knowledge to improve/develop

- IVD tests based on existing technologies
- Antimicrobials

✓ Real-time epidemiology

✓ Metagenomics

- A discovery platform for biomarkers
- A present quality-control platform
- A future diagnostic platform ?

Reverse Vaccinology

OPEN ACCESS Freely available online

PLoS GENETICS

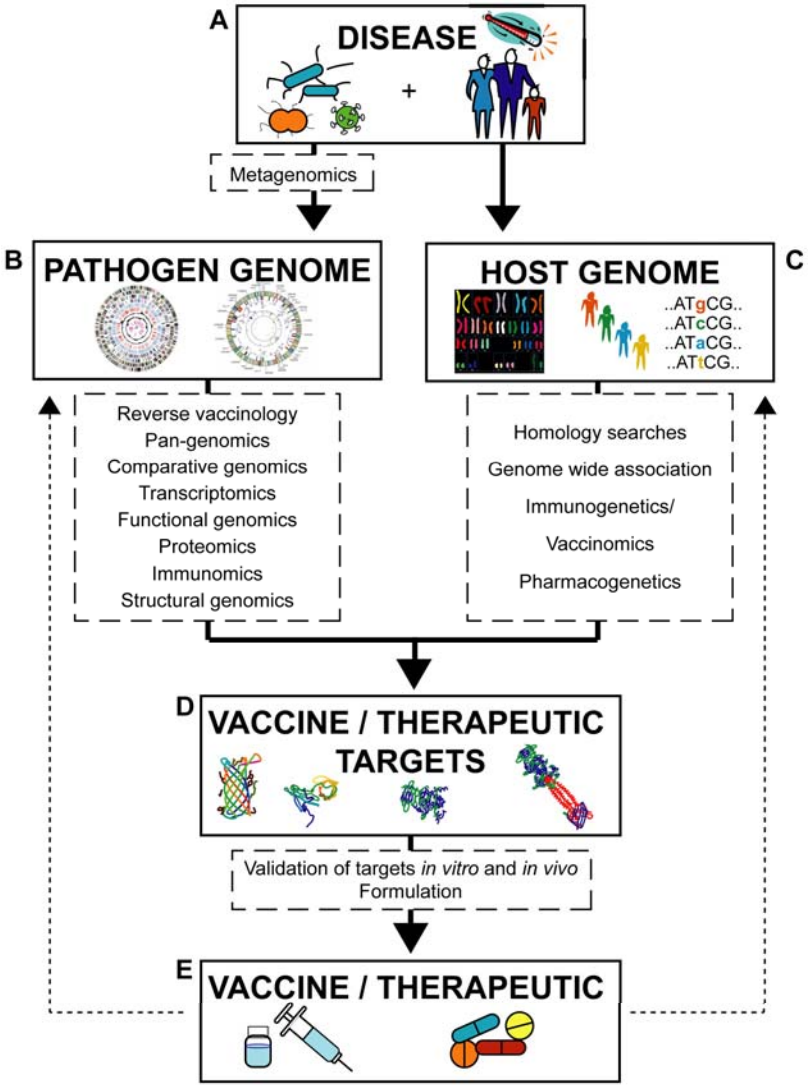
Review

The Key Role of Genomics in Modern Vaccine and Drug Design for Emerging Infectious Diseases

Kate L. Seib[1], Gordon Dougan[2], Rino Rappuoli[1]*

1 Novartis Vaccines and Diagnostics, Siena, Italy, 2 The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

- **Untargeted screening of patients to discover the causative agent**
- **In-silico vaccine and therapeutic targets discovery**
- **Validation of candidate vaccine and therapeutic targets**
- **Real-time monitoring of emerging pathogens**
  - ✓ Escape immune system in vaccinated subjects
  - ✓ Antibiotic resistance



Example of applications

➢ Infectious Diseases

- ✓ **Inject knowledge to improve/develop**
  - IVD tests based on existing technologies
  - Antimicrobials
- ✓ **Real-time epidemiology**
- ✓ **Metagenomics**
  - A discovery platform for biomarkers
  - A present quality-control platform
  - A future diagnostic platform ?

Leverage Real-time Epidemiology

MODERNISING MEDICAL MICROBIOLOGY

Skip to content | Accessibility | Sitemap

Search Site

UNIVERSITY OF OXFORD

HOME | PATHOGENS | OVERARCHING THEMES | STAFF | LINKS | PUBLICATIONS | JOBS | CONTACT

You are here: Home

UK CRC Modernising Medical Microbiology Consortium

UK wide consortium funded by the Wellcome Trust and MRC, between the University of Oxford, Health Protection Agency (HPA) and the Wellcome Trust Sanger Institute; to establish how revolutionary new technologies can be optimally intergrated into microbiology research and service strategic vision

MRC Medical Research Council

wellcome trust

Health Protection Agency

wellcome trust sanger institute

NHS National Institute for Health Research

Oxford Biomedical Research Centre

New Paradigm

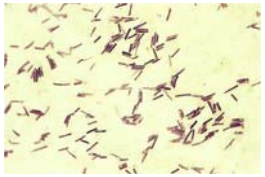
"**Translation of sequence-based typing from whole genome sequencing into clinical practice** via interrogation of databases by individual practitioners **will be a major behaviour change**"

Step 1 - Infection Surveillance

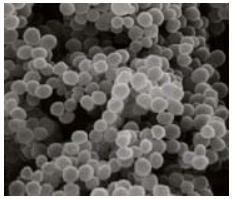
"One goal is to develop **web-accessible informatics tools** which are attractive and easy-to-use, **enabling local practitioners to direct their routine infection control practice efficiently**"

Step 2 - Association Study (Host-pathogen)

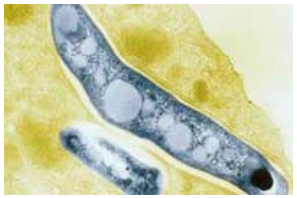
C. Difficile



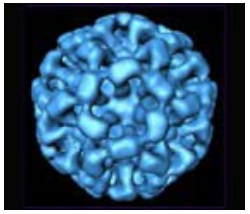
S. Aureus



M. Tuberculosis



Norovirus



Example of applications

➢ Infectious Diseases

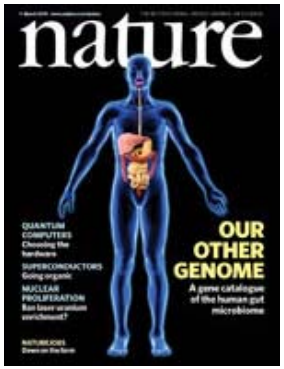
- ✓ **Inject knowledge to improve/develop**
  - IVD tests based on existing technologies
  - Antimicrobials
- ✓ **Real-time epidemiology**
- ✓ **Metagenomics**
  - A discovery platform for biomarkers
  - A present quality-control platform
  - A future diagnostic platform ?

Metagenomics and Biomarkers

Vol 464|4 March 2010| doi:10.1038/nature08821 nature

ARTICLES

A human gut microbial gene catalogue established by metagenomic sequencing



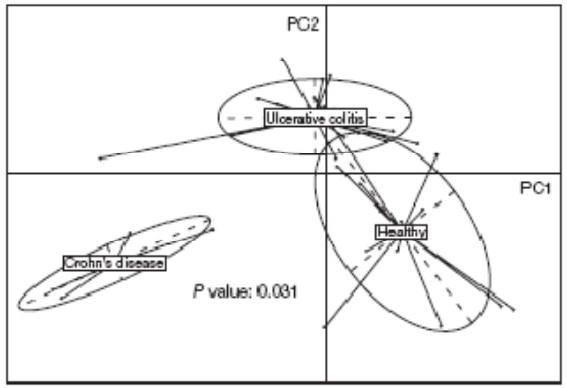


Figure 4 | Bacterial species abundance differentiates IBD patients and healthy individuals. Principal component analysis with health status as instrumental variables, based on the abundance of 155 species with ≥1% genome coverage by the Illumina reads in at least 1 individual of the cohort, was carried out with 14 healthy individuals and 25 IBD patients (21 ulcerative colitis and 4 Crohn's disease) from Spain (Supplementary Table 1). Two first components (PC1 and PC2) were plotted and represented 7.3% of whole inertia. Individuals (represented by points) were clustered and centre of gravity computed for each class; *P*-value of the link between health status and species abundance was assessed using a Monte-Carlo test (999 replicates).

- **Nutrisciences & Health**
  - ✓ Obesity
- **Pathology**
  - ✓ Inflammatory bowel disease

Example of applications

➢ Infectious Diseases

- ✓ **Inject knowledge to improve/develop**
  - IVD tests based on existing technologies
  - Antimicrobials
- ✓ **Real-time epidemiology**
- ✓ **Metagenomics**
  - A discovery platform for biomarkers
  - A present quality-control platform
  - A future diagnostic platform ?

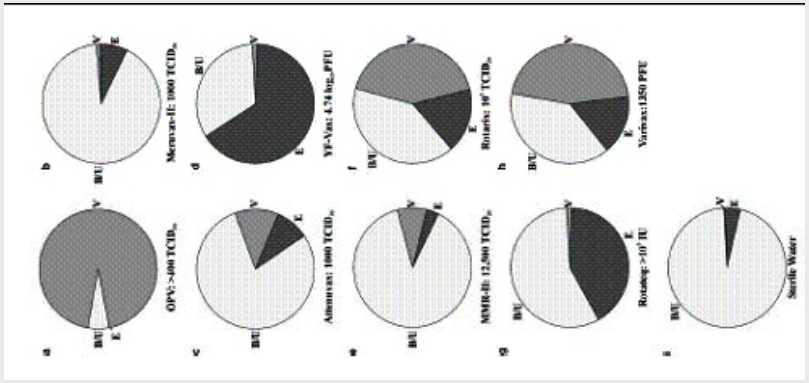
Quality Control of Viral Vaccines

Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus.

Joseph G. Victoria[1,2], Chunlin Wang[3], Morris S. Jones[4], Crystal Jaing[5], Kevin McLoughlin[5], Shea Gardner[5], Eric L. Delwart[1,2*]

Viral Metagenomics

RNA Extraction → cDNA + Random Amplification



QC of Viral Vaccines: pushed by FDA !

Vitrology News

Beckman Coulter Genomics and Vitrology Ltd. Partner to Offer Comprehensive Biologics Testing Solutions

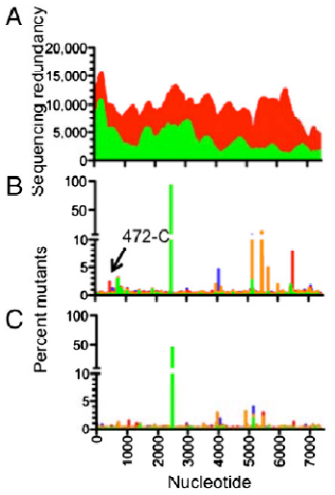


Fig. 1. MPS analysis of two batches of type 3 OPV performed by pyrose quencing. (*A*) The number of times each nucleotide was read in forwarc (green) and reverse (red) orientations. (*B* and *C*) Mutational profiles fo. vaccine batches that failed and passed the MNVT, respectively. Here and in all other figures the contents of mutants is shown by colored bars: mutations to A shown in orange, mutations to C in red, mutations to G in blue, and mutations to U in green.

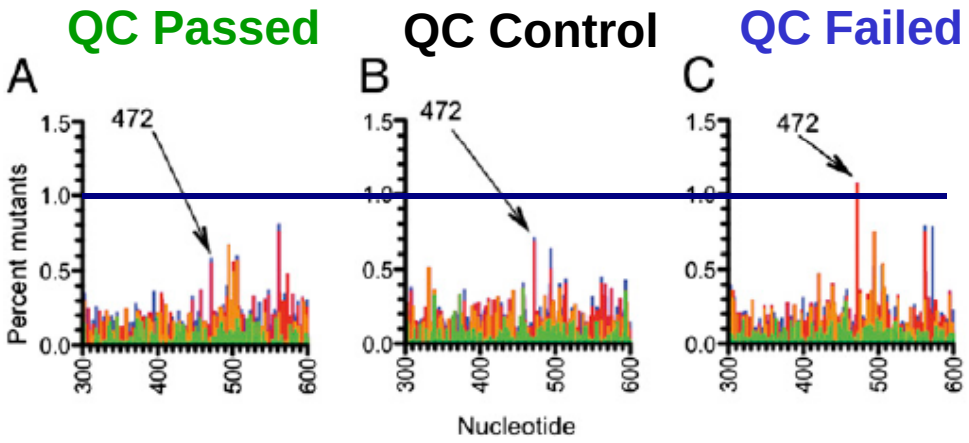


Fig. 3. Patterns of mutations in the vicinity of nucleotide 472 in Sabin 3 genome revealed by pyrosequencing. "Passed" WHO reference for MAPREC 96/572 (*A*); US National neurovirulence reference for type 3 OPV NC2 (*B*), and "failed" WHO reference for MAPREC 96/578 (*C*).

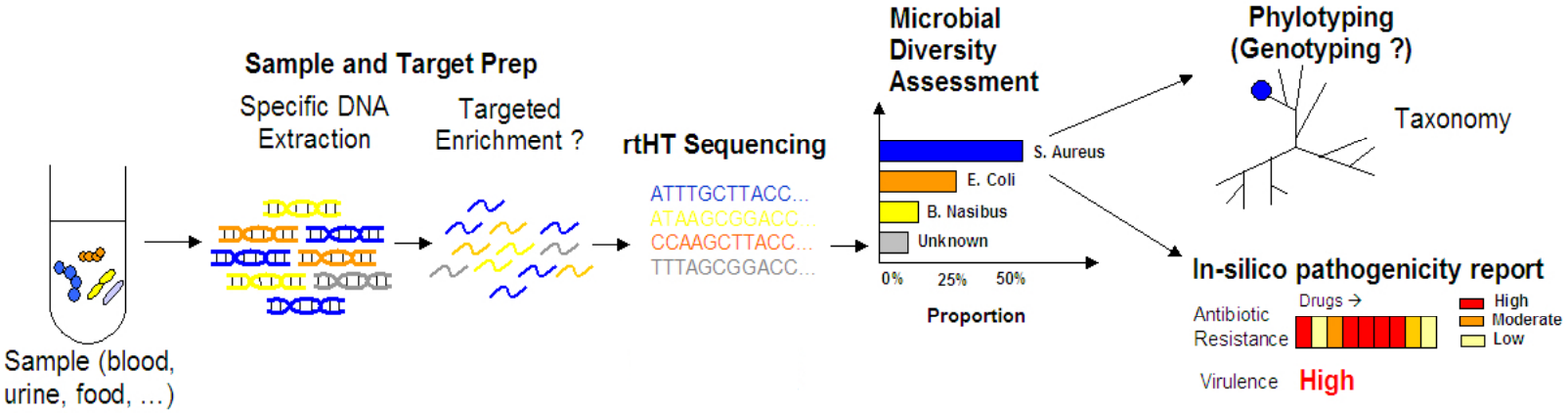
Example of applications

➢ Infectious Diseases

- ✓ **Inject knowledge to improve/develop**
  - IVD tests based on existing technologies
  - Antimicrobials
- ✓ **Real-time epidemiology**
- ✓ **Metagenomics**
  -
  - A present quality-control platform
  - A future diagnostic platform ?

Clinical Metagenomics

- **Direct sequencing from specimen**



- **Benefits**
  - ✓ **No need of culture (time consuming, bias)**
  - ✓ **Deliver simultaneously**
    - Identification/Typing (mixture)
    - Pathogenicity profile
  - ✓ **Platform versatility**
    - (m)DNA / RNA

Conclusion

- **We are living a revolution**
  - ✓ **Quantitative genomics: don't dream about (epi)genomic data, just get it !**
    - Rather dream about system biology

- **HT sequencing and clinical / industrial applications**
  - ✓ **Highly promising technology**
    - Content provider
    - Analyzer
  - ✓ **Current bottlenecks**
    - Pre/Post-analytics
    - Clinical validation
      - Large and well characterized cohorts

OPEN ACCESS Freely available online PLOS GENETICS

Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits

Nicholas Eriksson[1]*, J. Michael Macpherson[1], Joyce Y. Tung[1], Lawrence S. Hon[1], Brian Naughton[1], Serge Saxonov[1], Linda Avey[1], Anne Wojcicki[1], Itsik Pe'er[2], Joanna Mountain[1,3]*

1 23andMe, Mountain View, California, United States of America, 2 Department of Computer Science, Columbia University, New York, New York, United States of America, 3 Department of Anthropology, Stanford University, Stanford, California, United States of America