Analysis of Deep Sequencing Data to Study Tumor Biology

UseR! 2009, Rennes, July 2009

Wolfgang.Raffelsberger@igbmc.fr

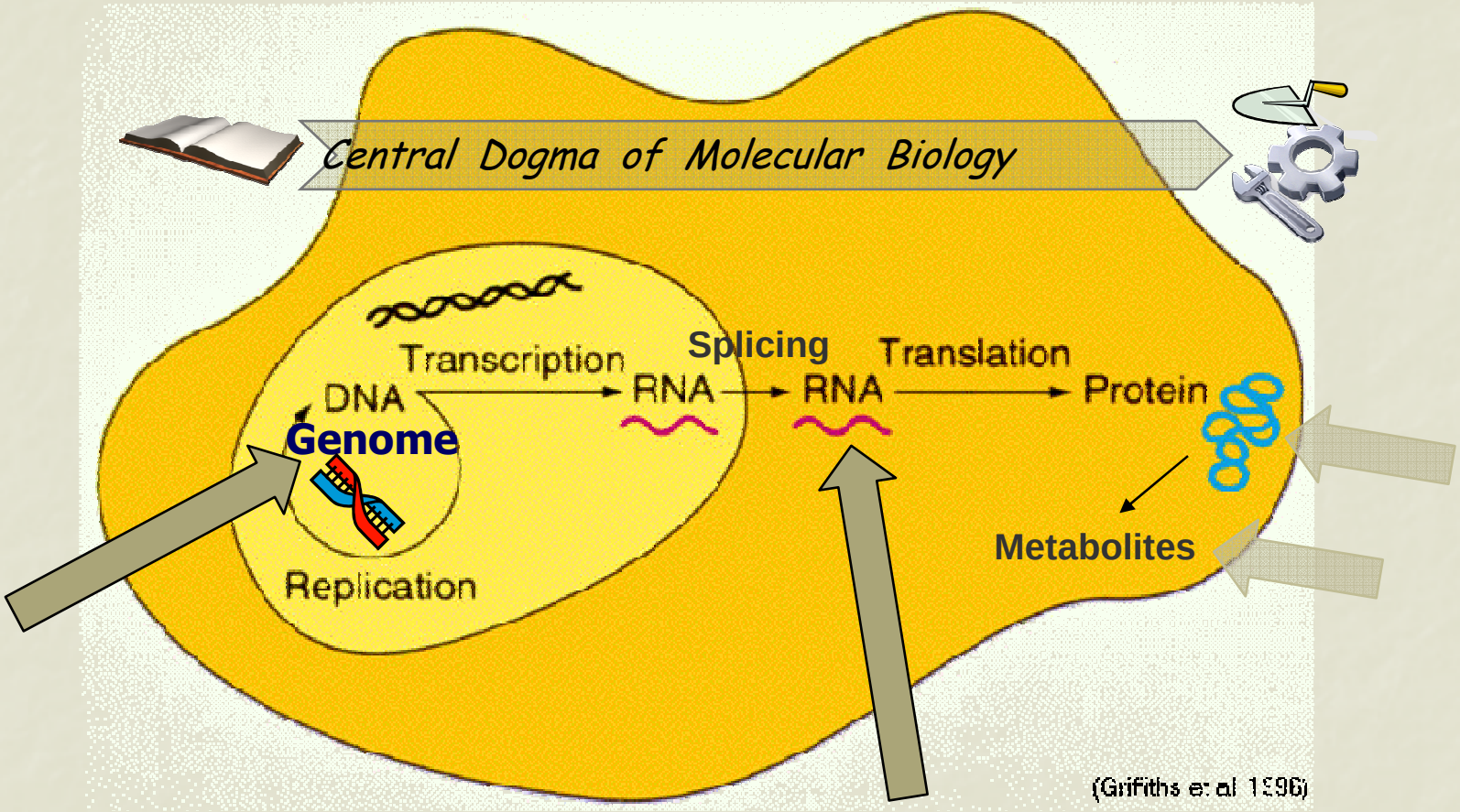
Nocodème Paul, Olivier Poch

LBGI

Laboratoire de **B**io-Informatique et **G**énomique **I**ntégratives

IGBMC, Strasbourg

Omics Measures



Cancer : Genome abnormalities in transformed cells

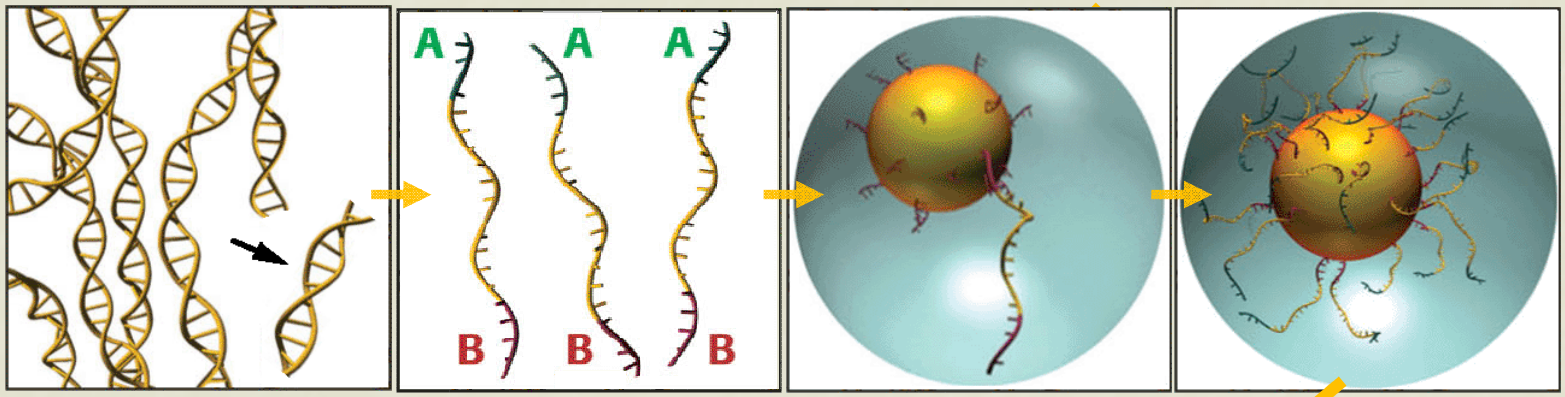
- **Carcinogens → abnormalities** (tobacco, radiation, chemicals, infectious agents)
- **Randomly acquired** through errors in DNA replication (+ DNA methylation, microRNAs)

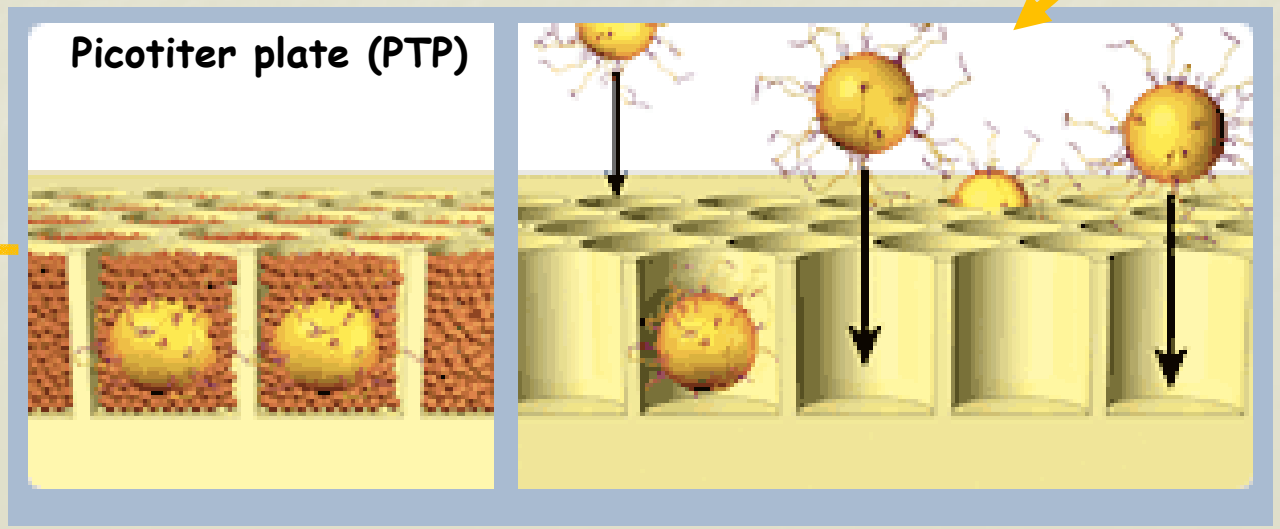
Roche 454 Pyrosequencing

www.454.com

5-100ng DNA → Amplification :
- whole genome *or*
- specific fragments

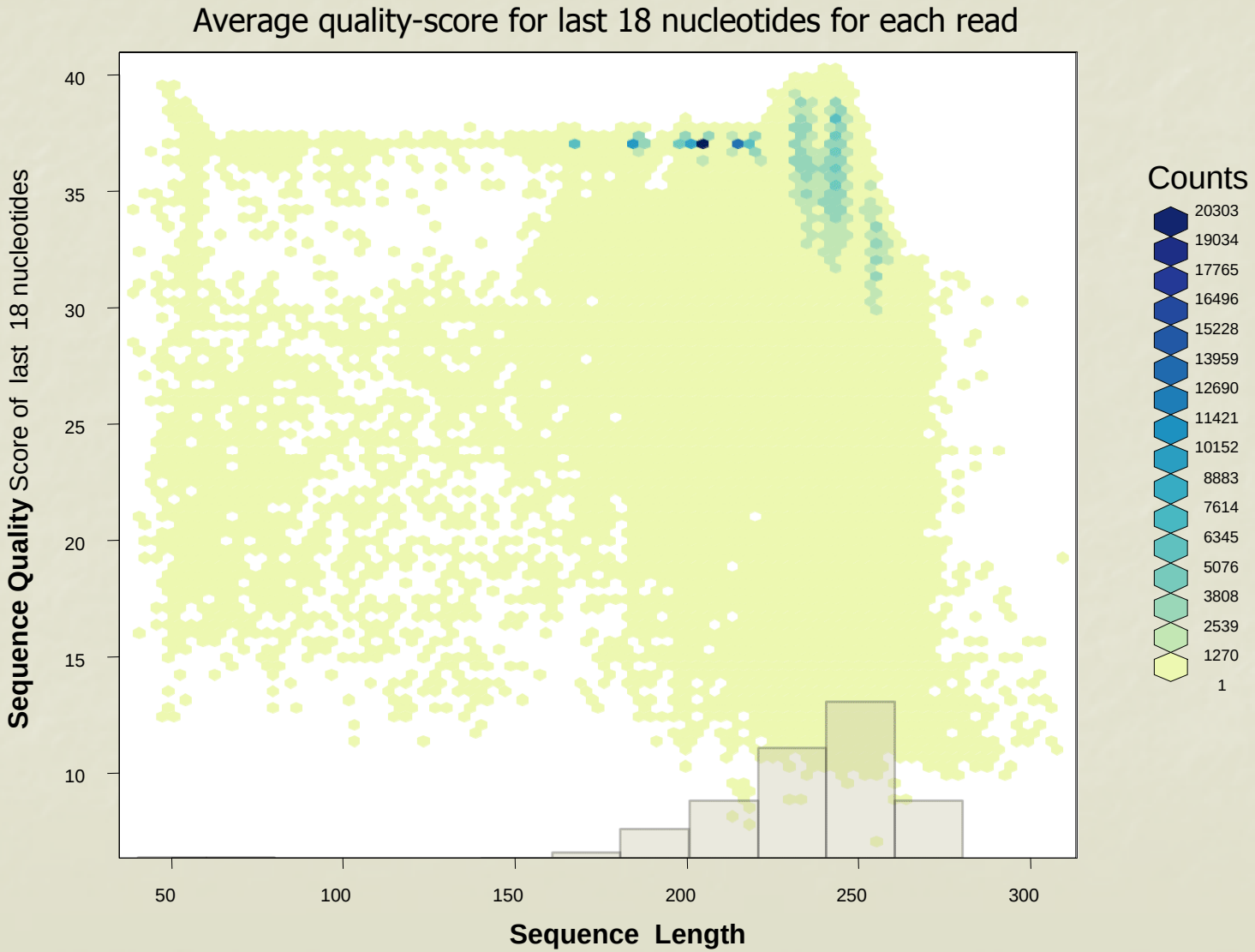
Emulsion PCR :





Sequencing by Synthesis

Relation between Read-Length and Sequence Score

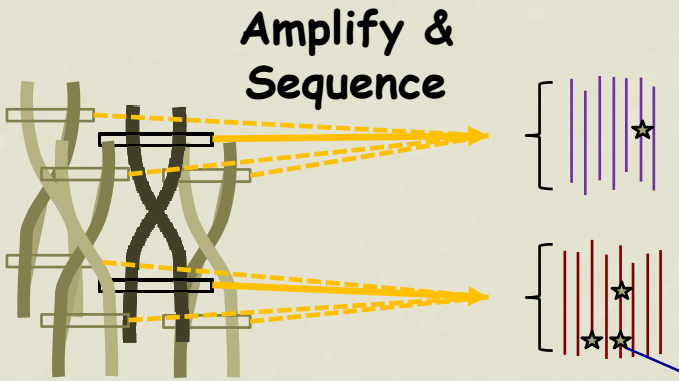


Experiment : Selected Genome Regions

Pilot-study : Study selected regions on genome

High fidelity PCR protocol of multiple genes

Single Patient : Population of tumor cells



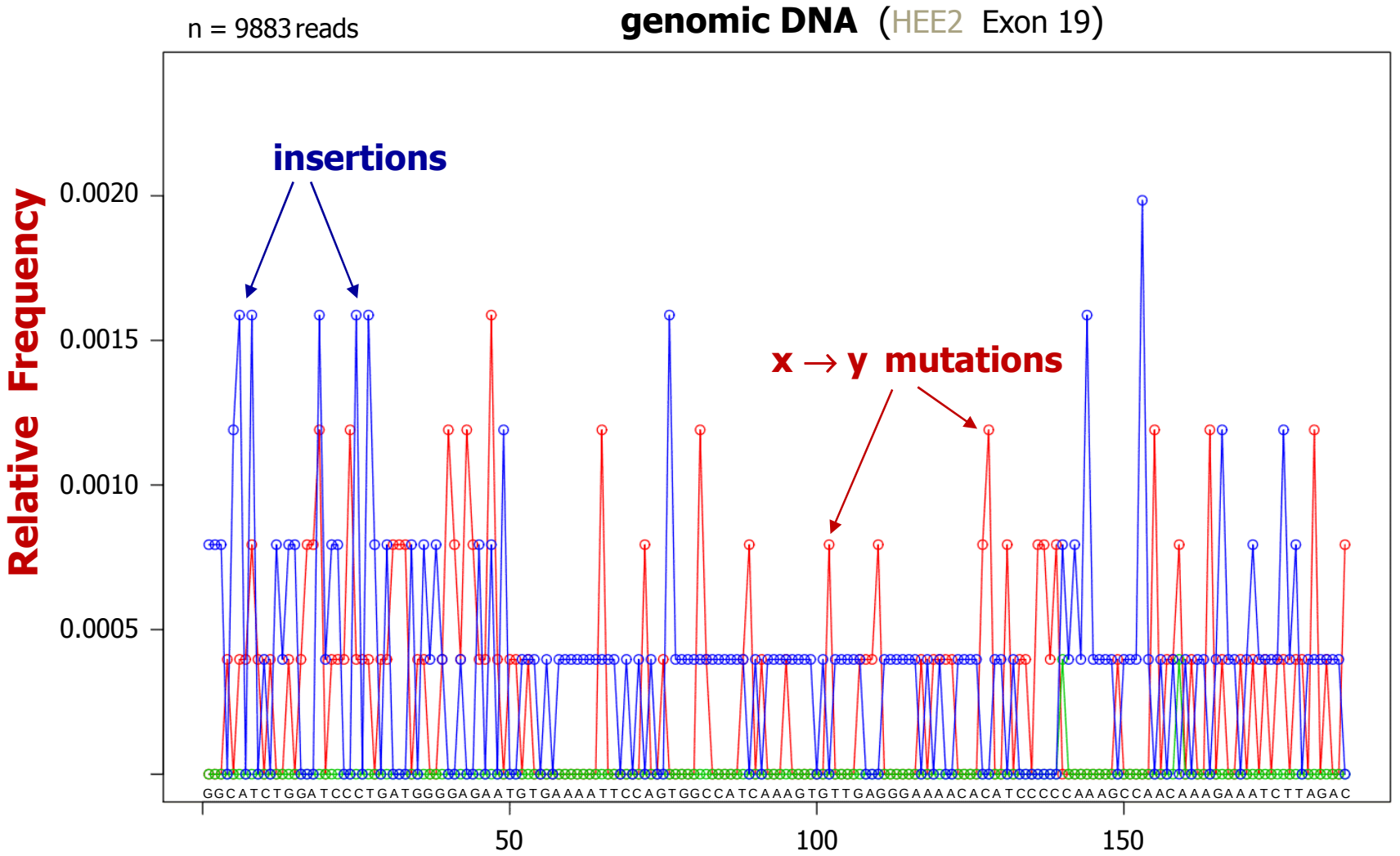
(genomic DNA :)

selected Exons + flanking intronic regions

Sequence alterations

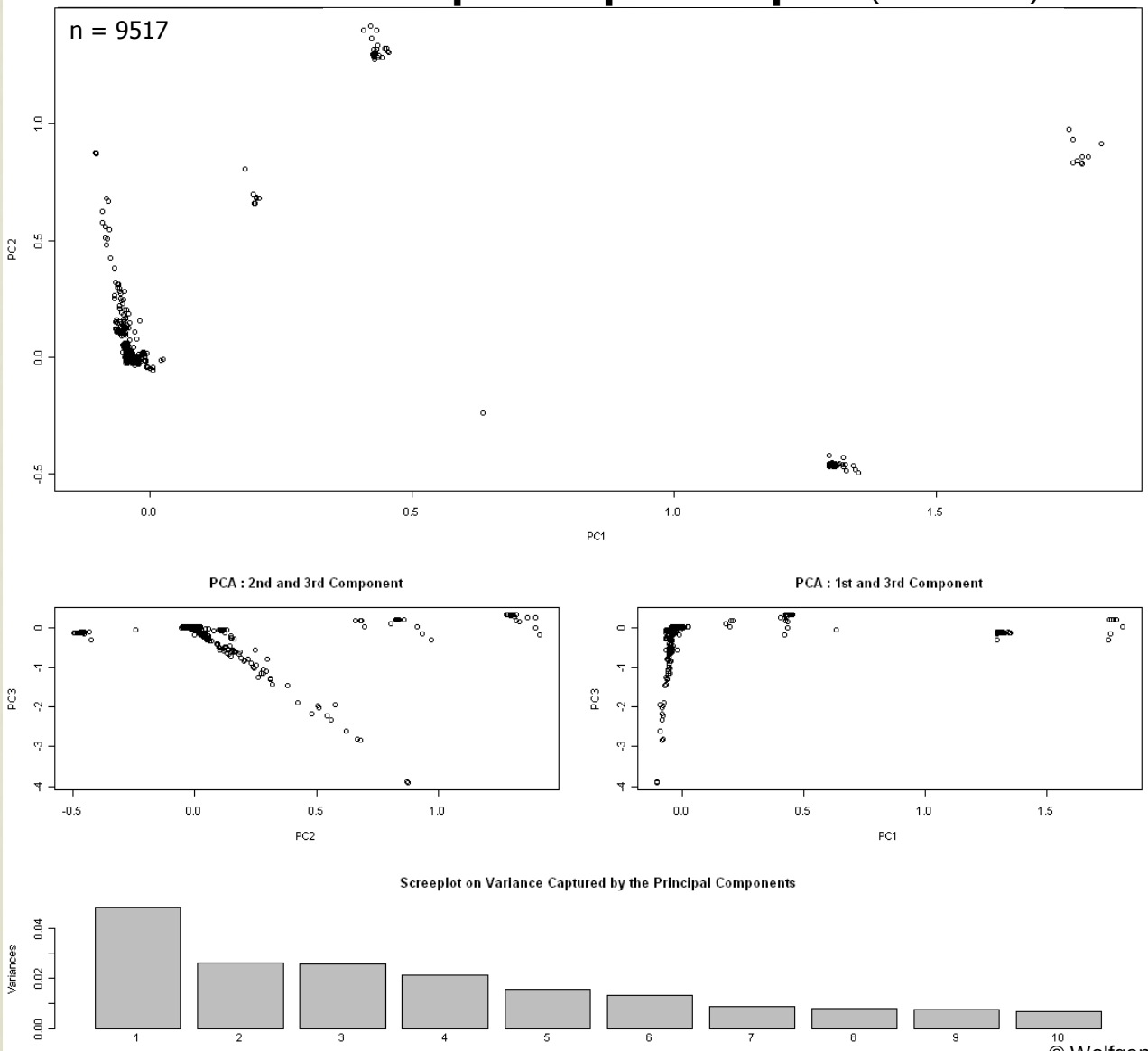
⇒ **Tumor Clonality**

Lung Cancer Deep Sequencing (454)



Lung Cancer Deep Sequencing (454)

Principal components plot (HEE exon 20)



Combining Transcriptome *AND* Genome Deep Sequencing

Pilot-study : Study selected regions on genome *and* mRNA

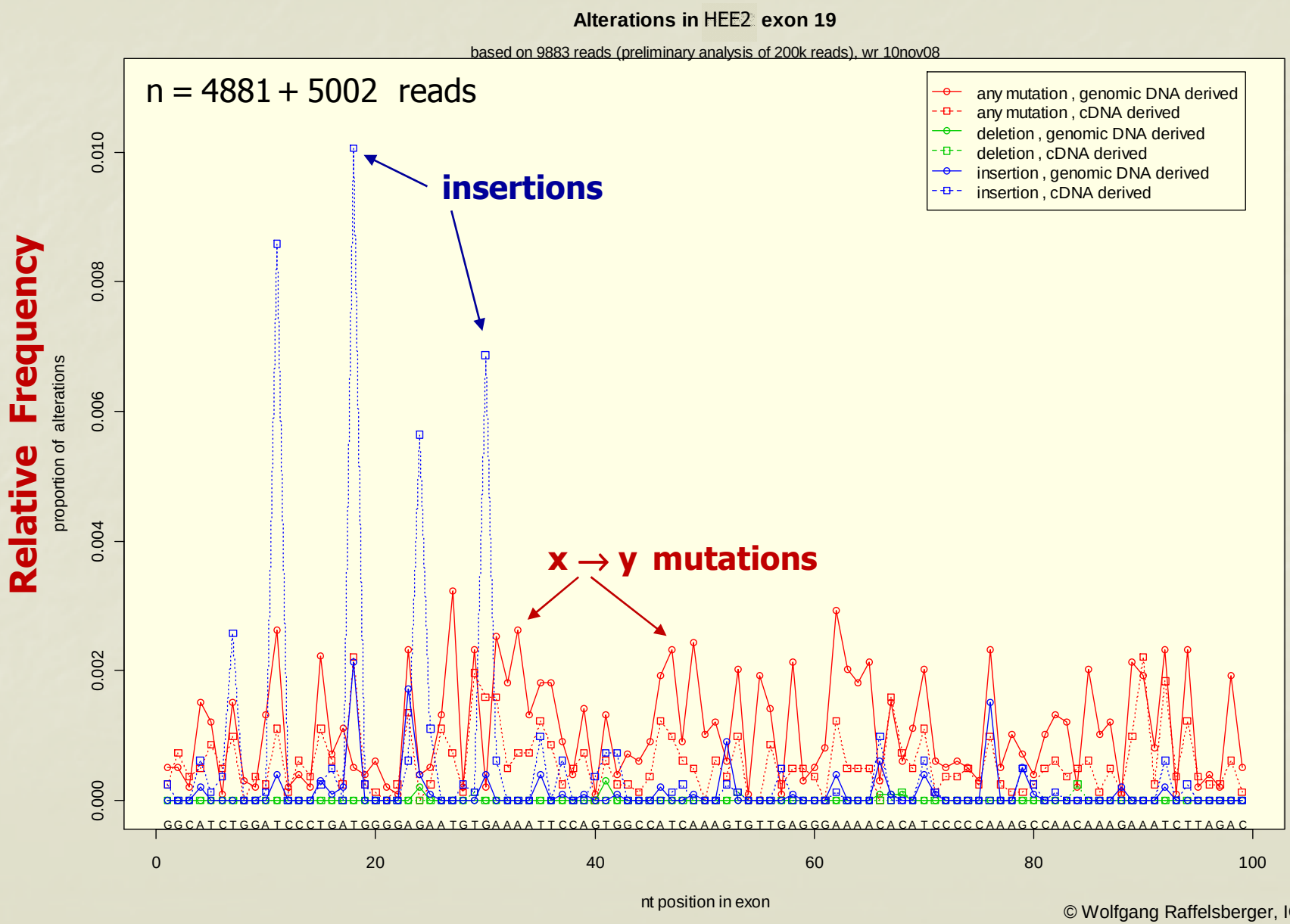
High fidelity PCR protocol of multiple genes :

- ***genomic DNA***
  selected Exons + flanking intronic regions

- ***mRNA***
  selected Exons *with* splice boundaries

⇒ Transcriptional Infidelity

Lung Cancer Deep Sequencing (454)



Challenge : Avoid Sequencing Artifacts ?

DNA - Alteration : *(Mutation, InDel)*

- **Frequency on sense & anti-sense strand**

  Expect same frequency for true alterations

- **Site : compare with sequence quality cores**

  Low Sequence Score : increased likelihood of artifacts

- **Typical sequence pattern in vicinity of sites ?**

  May have biological context …

Tools

- **Biostrings** (pairwise alignments …)
- **ShortRead**
- **rareSNPtools** *(to be released soon)*

  - (further) interpreting pair-wise alignment results
  - reduce In/Del complexity to matrix-like representation
  - SNP counting & localization
  - plotting
  - (in progress:) integration of Phred-Score
    monitor impact on protein mutations
    likelihood of sequencing artifacts

Conclusions Deep Sequencing

New options to study genome *and* mRNA

Novel challenges due to wealth of information

- Universal platform to combine multiple aspects (?)
- Specialized methods still in development
- Expert interventions
- Large scale biological interpretation *(references ?)*
- Cost & infrastructure

Acknowledgments

IGBMC

Computing Facility S Uge, G Seith

Microarray Platform C Thibault-Charpentier

Platform TCA L Brino

Equipe Oudet/Schultz J Papillon

LBGI & **Plate-forme Bioinformatique de Strasbourg**

CANCÉROPÔLE DU GRAND-EST

Hôpital Hautepierre, CGE
E Guerin, M Guenneugues,
MP Gaub, M Beau-Faller, P Oudet

Inserm **U682** C Nicolet, D Guenot

dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

cea

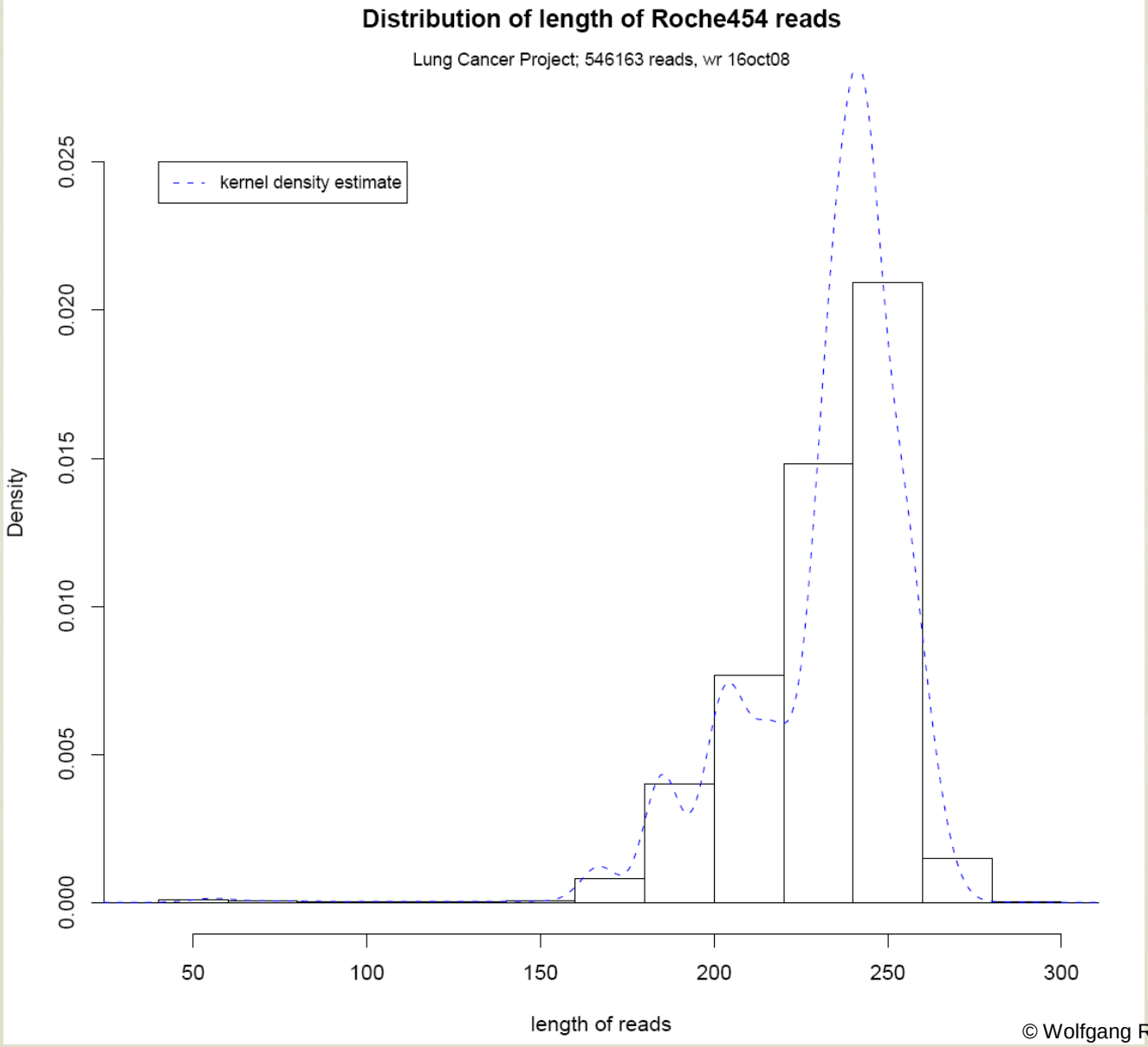
Loria

Université de Montréal

Oncotest GmbH Prof. H.H. Fiebig

Inserm cnrs UNIVERSITÉ DE STRASBOURG

Lung Cancer Deep Sequencing (454)



Transcriptomics Perspectives with Deep Sequencing

- **Unbiased** search (unexpected transcripts …)
- Measure **expression *and*** detect **polymorphisms**
- Clear distinction of signal *above* **background**
  occurrence of sequences : simple and intuitive count
- Capacity to detect **splicing** variants :
  - requires rather high sequencing coverage
  - analysis **complex** ← *short read - length !!*
    ↖ *high needs for expert intervention*
- More uniform data for meta-analysis ?

Cancer

Cancer (= malignant neoplasm)

Group of cells display uncontrolled growth :

- **Division beyond normal limits**
- **Invasion** (intrusion, destruction of adjacent cells)
- **Metastasis** (spread to other locations)

different to benign tumors

- Malignant characteristics vary in **in-homogenous manner**
- **Causes 13% of all deaths**

Genome abnormalities in transformed cells

- **Carcinogens → abnormalities** (tobacco, radiation, chemicals, infectious agents)
- **Other** cancer-promoting **genetic abnormalities :**
  randomly acquired through **errors in DNA replication**
  heritability : complex interactions between carcinogens and host genome
- **New aspects** of the genetics of cancer pathogenesis :
  DNA methylation, microRNAs are increasingly important