

"-omics"

Based on their chemical properties, biological molecules can be fractionated in several classes : nucleic acids (DNA, RNA) , proteins, metabolites, lipids, glucides,

Although large, the complexity of each class within an organism is limited.

For humans

Diploid genome	: 6 10^9 base pairs
Transcriptome	: 23 000 genes $\approx 10^5$ transcripts
Proteome	: $\approx 10^6$ proteins
Metabolome	: several hundred/thousand
Etc ...	

For almost each class , new technologies enable the exploration of this complexity as a whole :

DNA => genomics

RNA => transcriptomics

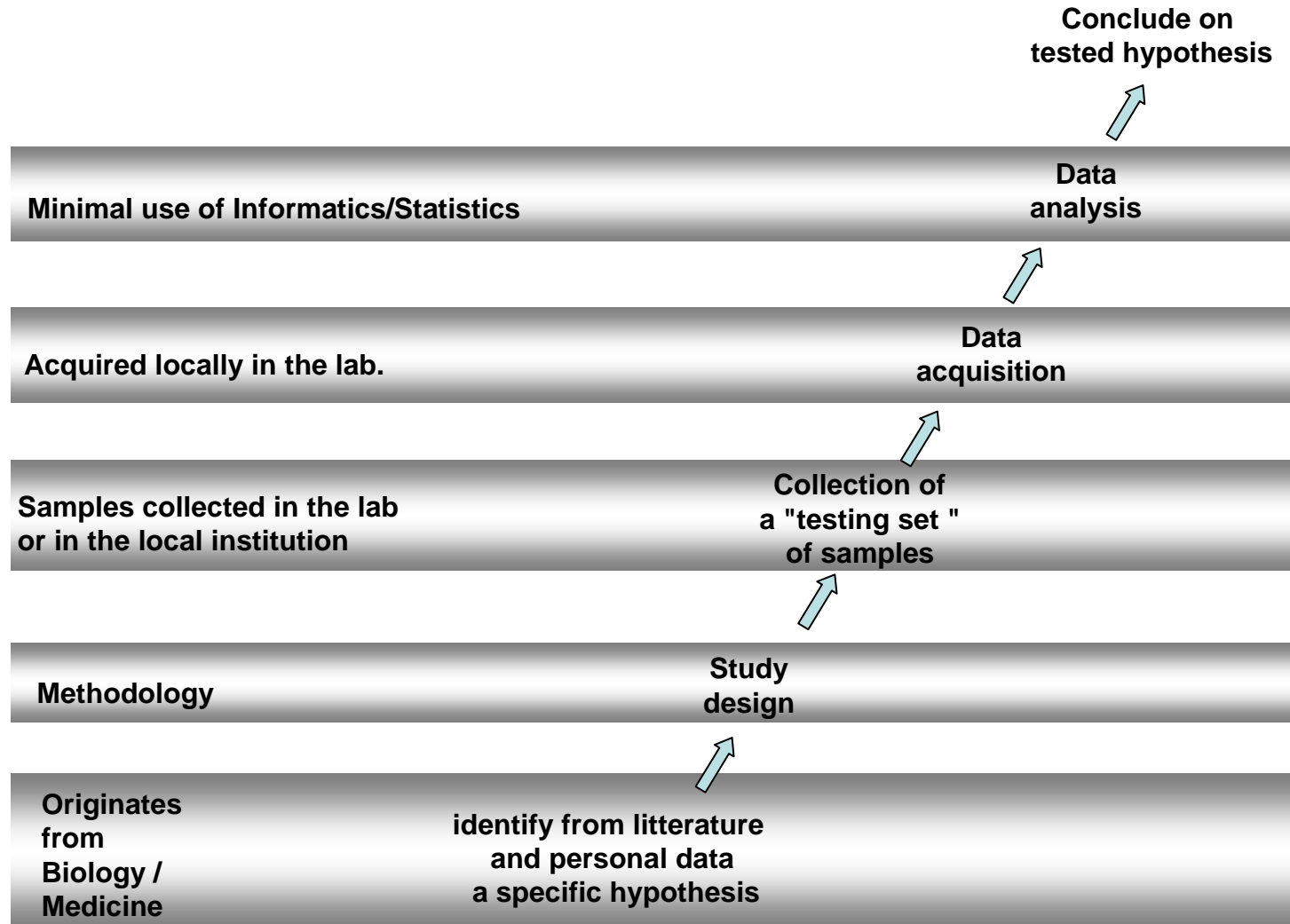
protein => proteomics

metabolites => metabolomics

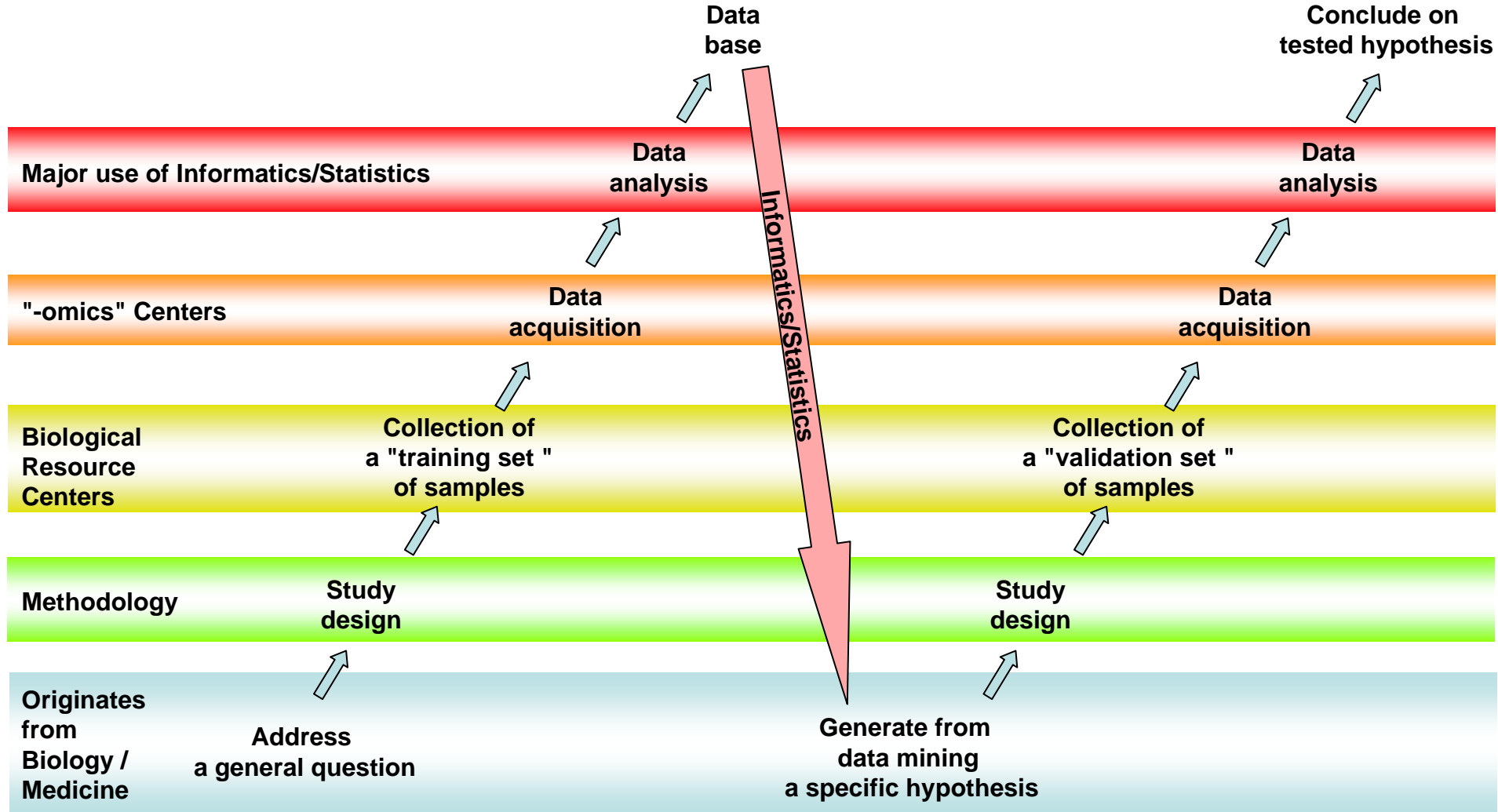
Etc ...

The classic workflow in research:

Mostly a one team endeavor in a single institution



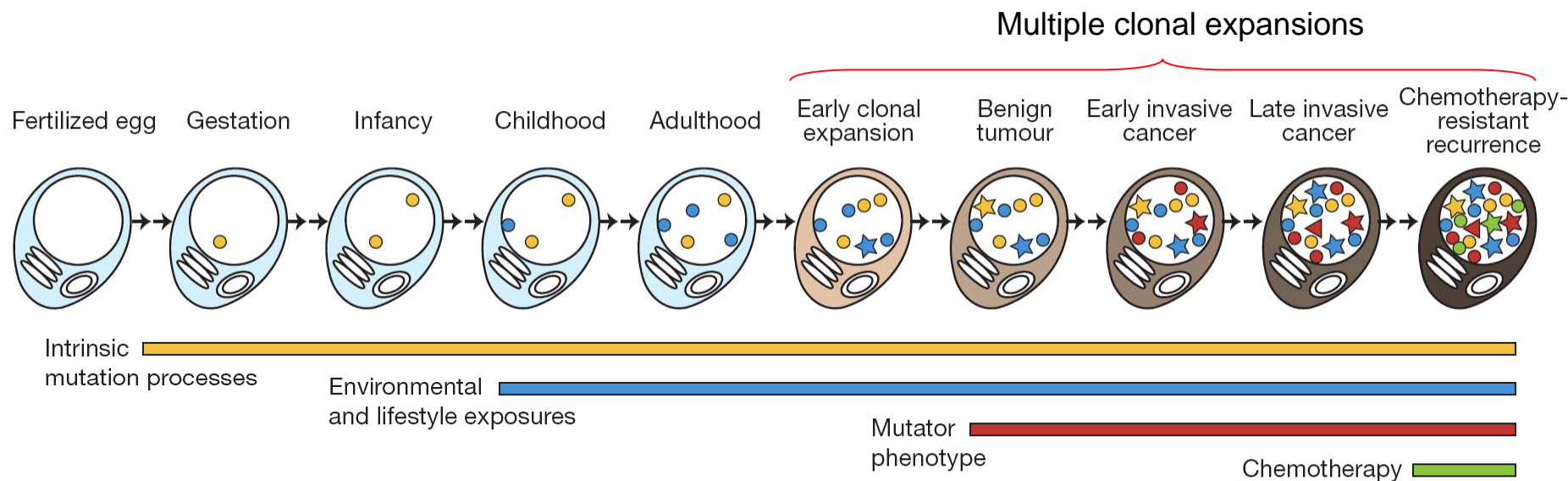
The "-omics" workflow in research: The emergence of team research



Among the different -omics, genomics and transcriptomics have a special place

- Genomics and transcriptomics address the same class of biomolecules :
the nucleic acids.
 - Simple (linear), highly informative structure.
 - Similar chemical properties that can be effectively used for automation
(RNA frequently studied after retrotranscription to DNA).
 - Similar analytical procedures (electrophoresis, hybridization, sequencing, ...) ; same machines,
- Both have to date the best implemented methods ;
both keep driving the -omic revolution.
- Next generation sequencing is a major breakthrough :
 - ultra high throughput data generation
 - performed on stable molecules (DNA)
 - soon performed on single molecules
 - provide digital data (large dynamic range)
 - can be further automated
 - will be very cheap (economy of scale achieved by large centers)

Somatic genetic events from the fertilized egg to an aggressive cancer cell



Classification of somatic genetic events :

Structure : "point" mutations , translocation, deletion, amplification

Function : "passengers" versus "drivers"

Traditional strategies to identify candidate "cancer" genes

Retrovirology

- transduced cellular genes
- site of viral insertion

Transfection assays

- NIH 3T3
- Embryonal fibroblasts

Cytogenetics

- balanced translocation
- deletion
- amplification

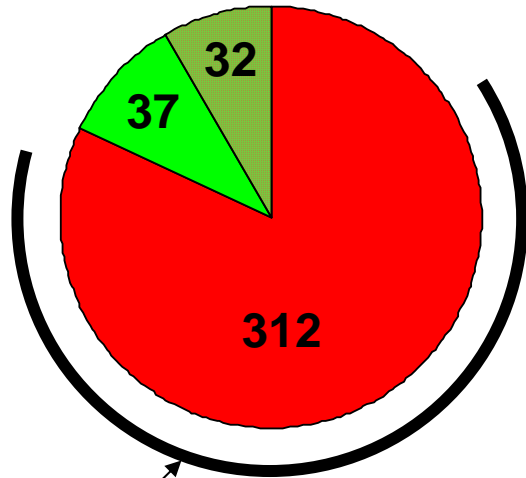
Positional cloning of high penetrance cancer susceptibility genes



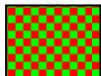
Functional candidate

- pathways
- gene families

June 2009

About 380 "cancer" genes known for humans



-  Somatic mutations in cancer
-  Germline mutations predisposing to cancer
-  Both somatic and germline mutations

295 identified from chromosome translocation in hemopathies and sarcomas

Today, many routinely implemented tests for cancer patients are based on the identification of DNA mutations/translocations/amplifications

Cancer prevention and early detection

High penetrance predisposition to cancer
(BRCA1&2, MSH2, MLH1, MSH6,)

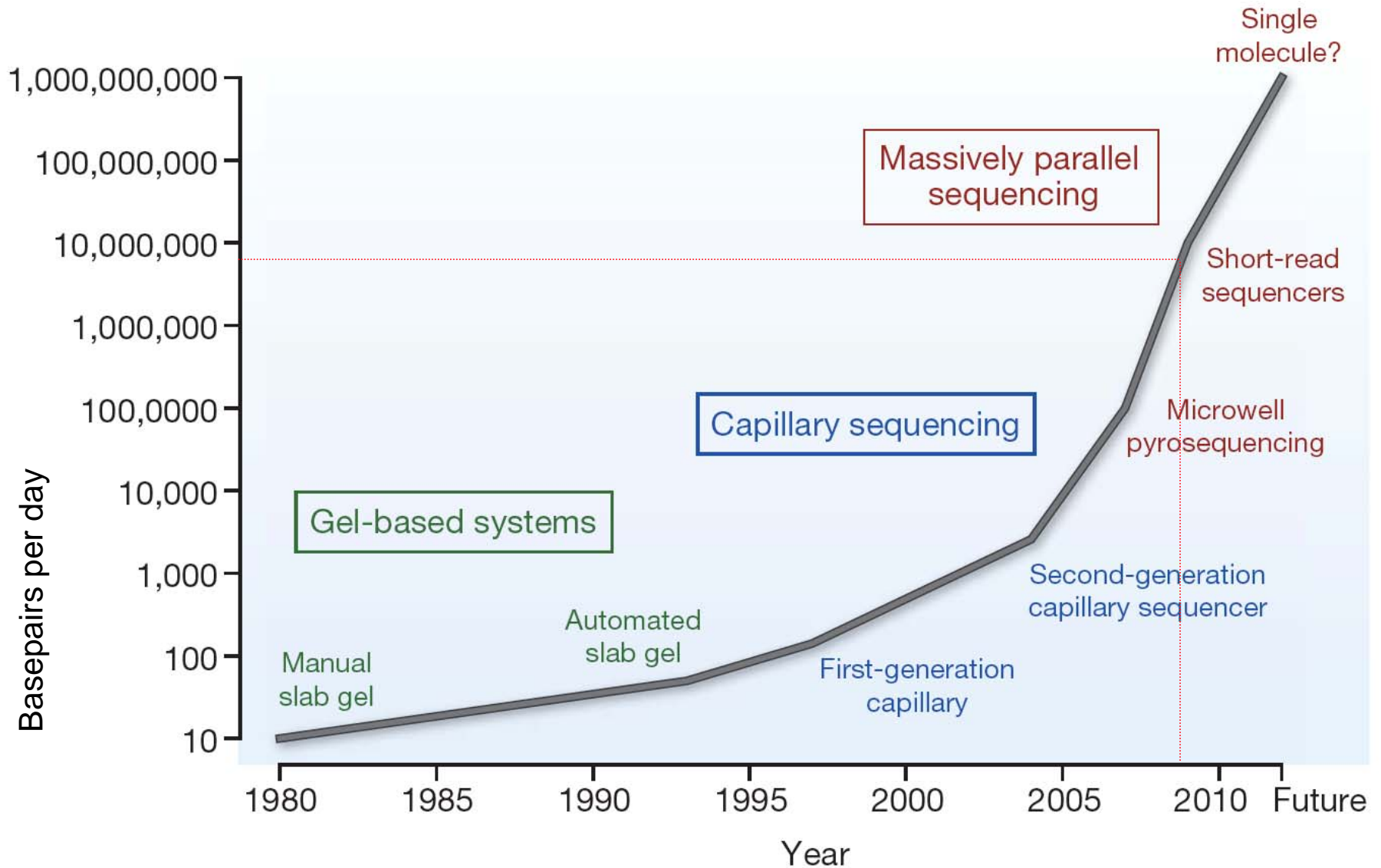
Diagnosis (and prognosis)

- Many fusion genes mostly in hematological diseases and sarcomas.
- KIT mutation for mastocytosis
- KIT or PDGFRA mutations gastrointestinal stromal tumors
- JAK2 mutations for myeloproliferative disorder
- Somatic mutation rate in IGHV in grade3 CLL

Response to treatment

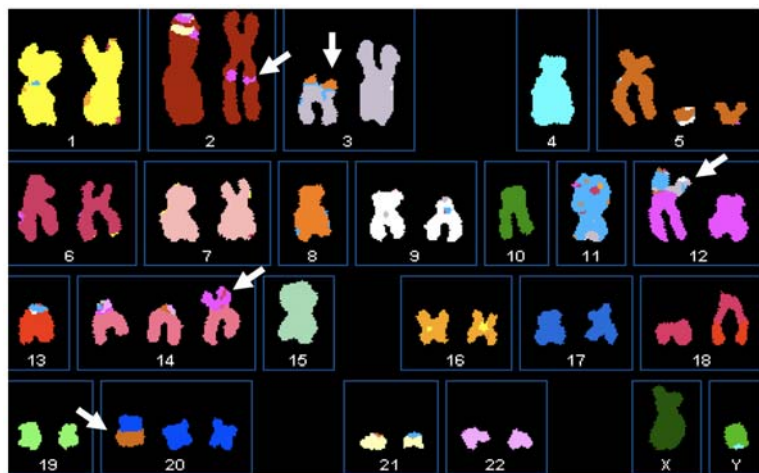
- HER2 amplification for Herceptin in breast cancer
- KRAS and BRAF mutations for Cetuximab in colorectal cancer
- EGFR mutation for Gefitinib in NSC lung cancer
- ABL mutations for Imatinib in CML
- KIT for Imatinib in some sarcomas and related tumors
- FLT3 , NPM1, CEBPA mutations in AML (evaluate interest of engraftment)
- MGMT methylation for Temolozomid in glioblastoma
- Early detection of relapse

Sequencing techniques over the last 30 years

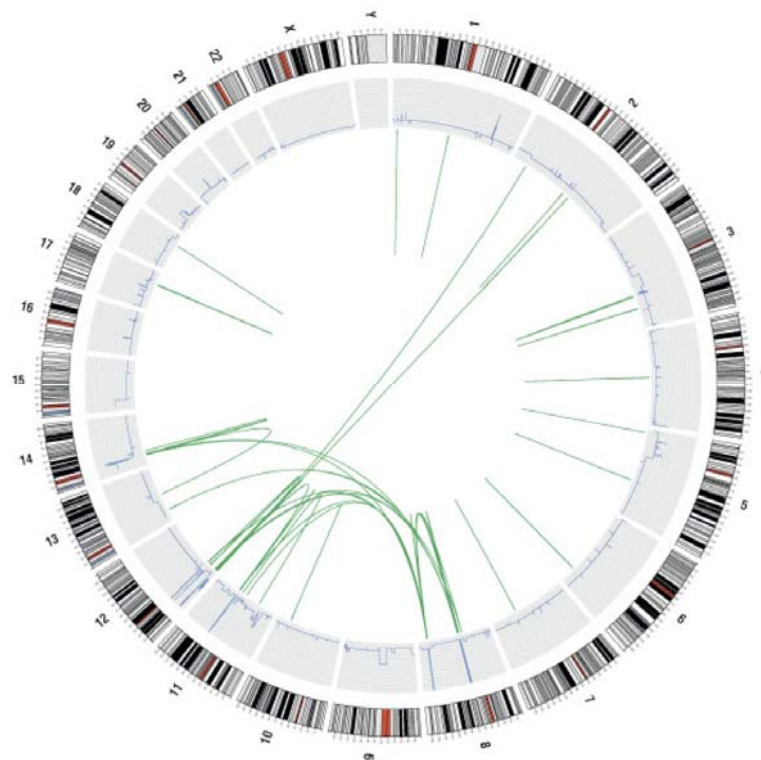


adapted from M. Stratton

Chromosome rearrangement in the small cell lung cancer NCI-H2171 cell line

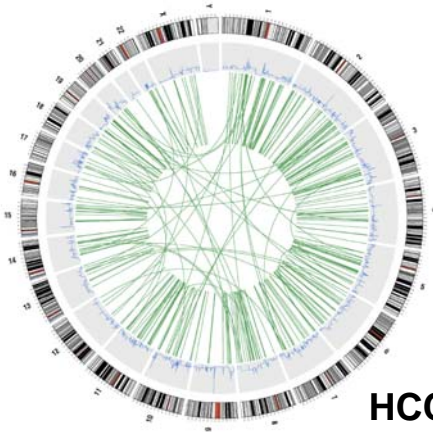


Spectral Karyotyping
~10 chromosome rearrangements
(all interchromosomal)

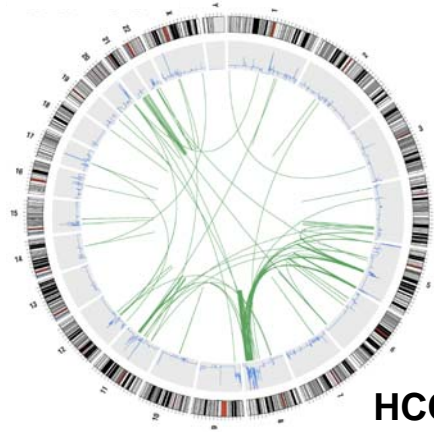


Paired end sequencing
88 chromosome rearrangements
(59 intrachromosomal, 22 interchromosomal)

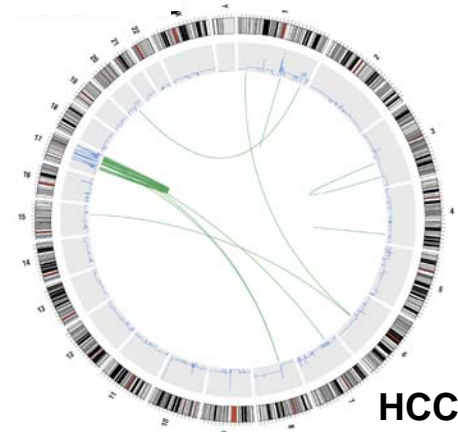
Chromosome rearrangements in breast cancer



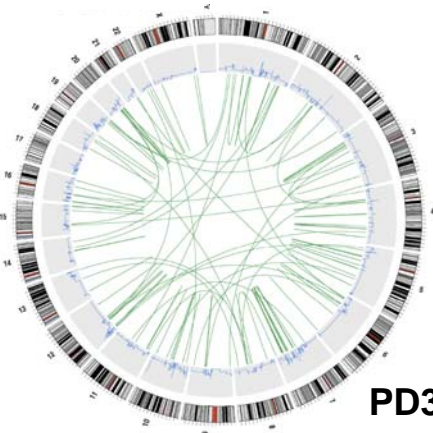
HCC38



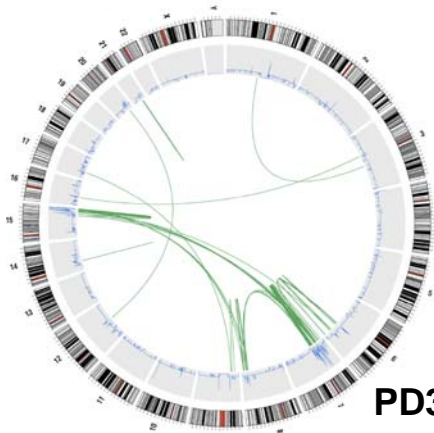
HCC1954



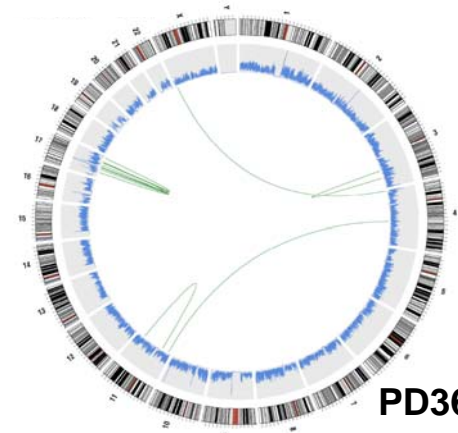
HCC2218



PD3665a

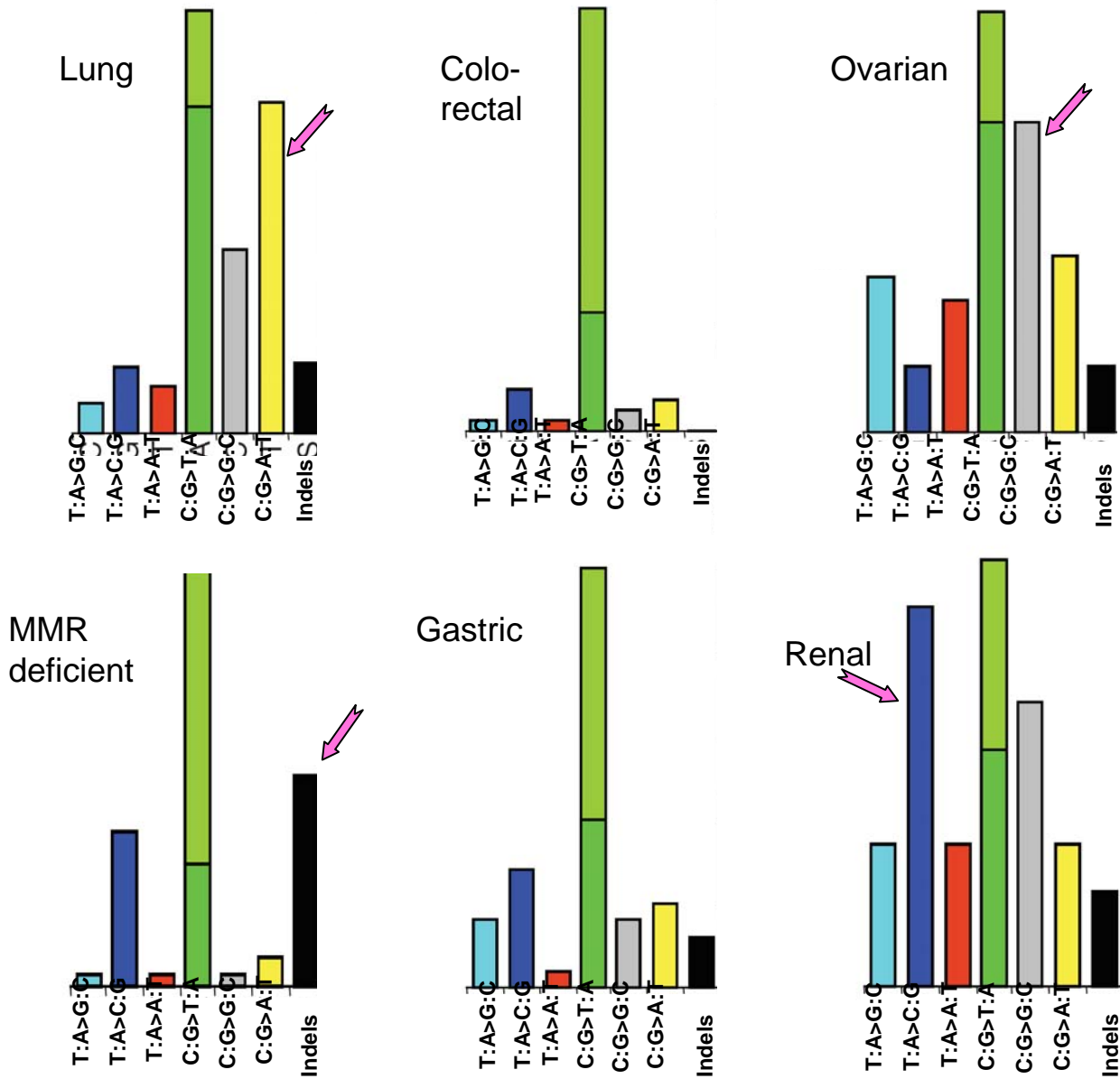


PD3668a



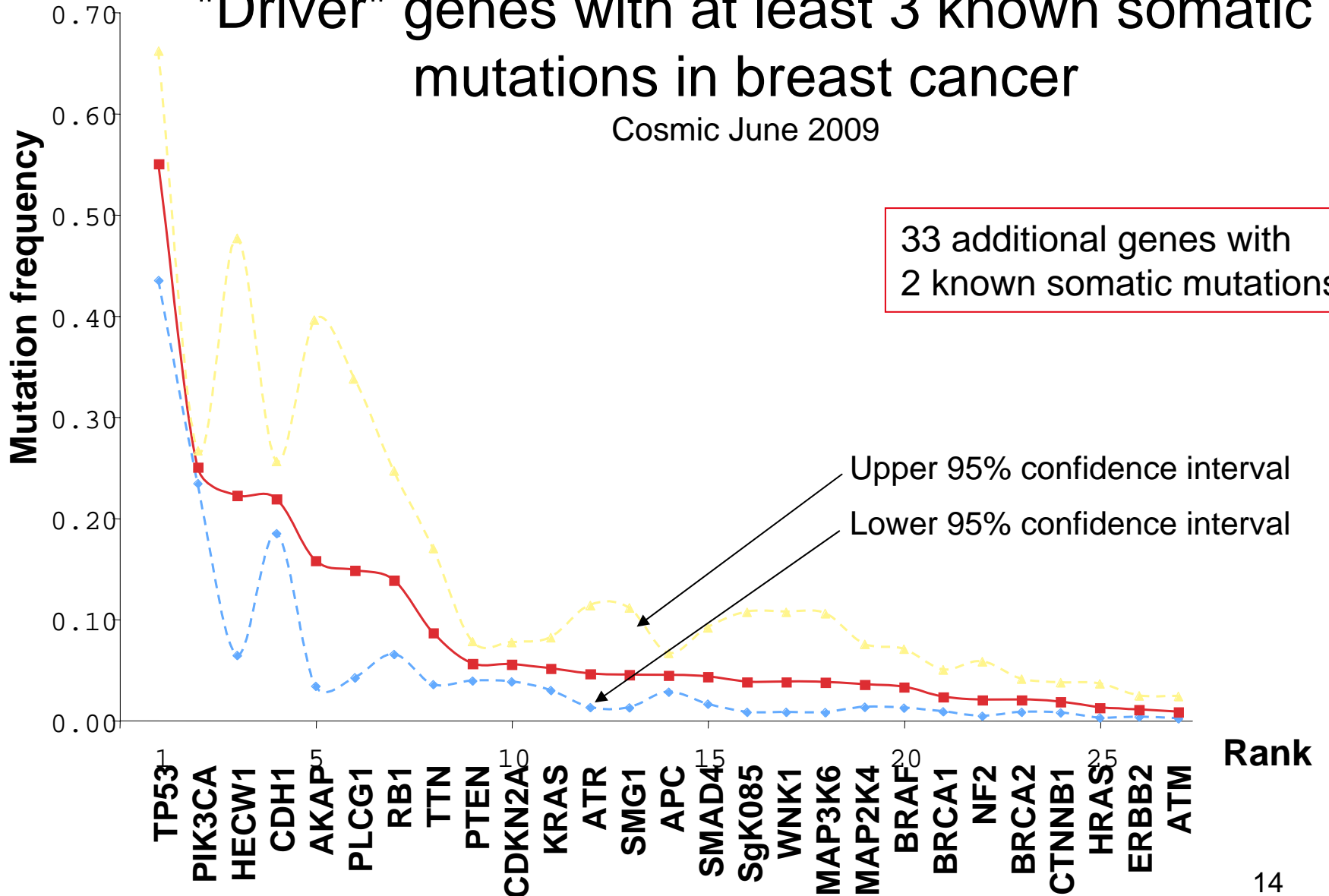
PD3671a

The spectrum of somatic mutations differs with tumor type



"Driver" genes with at least 3 known somatic mutations in breast cancer

Cosmic June 2009



A simplified view of somatic mutations in an individual tumor

- A large number (highly variable from $\approx 10^3$ to $\approx 10^5$) **passenger mutations** : analysis of the mutational spectrum provides information on tumor etiology and tumor progression

The knowledge of these mutations will provide information on cancer causation

- A smaller number of **driver mutations**, schematically subdivided into two sub-groups :

- one to 5 mutations distributed over a small set of frequently mutated genes (e.g. p53, PTEN, APC, ...)
- the rest distributed over a large set (50 genes?) of infrequently mutated genes

The knowledge of these mutations will determine patient management

Present applications of new generation sequencing in research

- **Full characterization of nucleic acids mixtures**
 - Metagenomics (complex combination of multiple microbial genomes)
 - Whole genome (mammalian)
 - Whole transcriptome (RNA-seq)
 - Whole methylome (epigenomes)
- **Targeted sequencing (combined with multiplexing)**
 - Long range PCR
 - Specific genes (e.g. BRCA1) or families of genes (e.g. tyrosine kinases)
 - Hybrid Capture (pull down):
 - All exons
 - All CpG islands
- **Chromatin conformation/interaction**
 - ChIP-seq

Medical sequencing

- In the near future, sequences from patient tumor and blood samples will be acquired for most cancer patients
- The derived information will be instrumental in fine diagnosis, selection of therapy and follow-up.
- However, it is presently unclear :
 - What proportion of the genome will be sequenced (from a few hundred genes to whole genome.)
 - What sequencing kind of sequencing facilities will be implemented (hospital labs, regional core facilities, private companies?)
 - How and where the sequences will be mined to derive information useful to patient management (bioinformatics , system biology)
 - What information will be transferred to the clinician/patient and will have to be kept in the patient medical file.
 - How clinician will be prepared to take the best advantage of this new information.