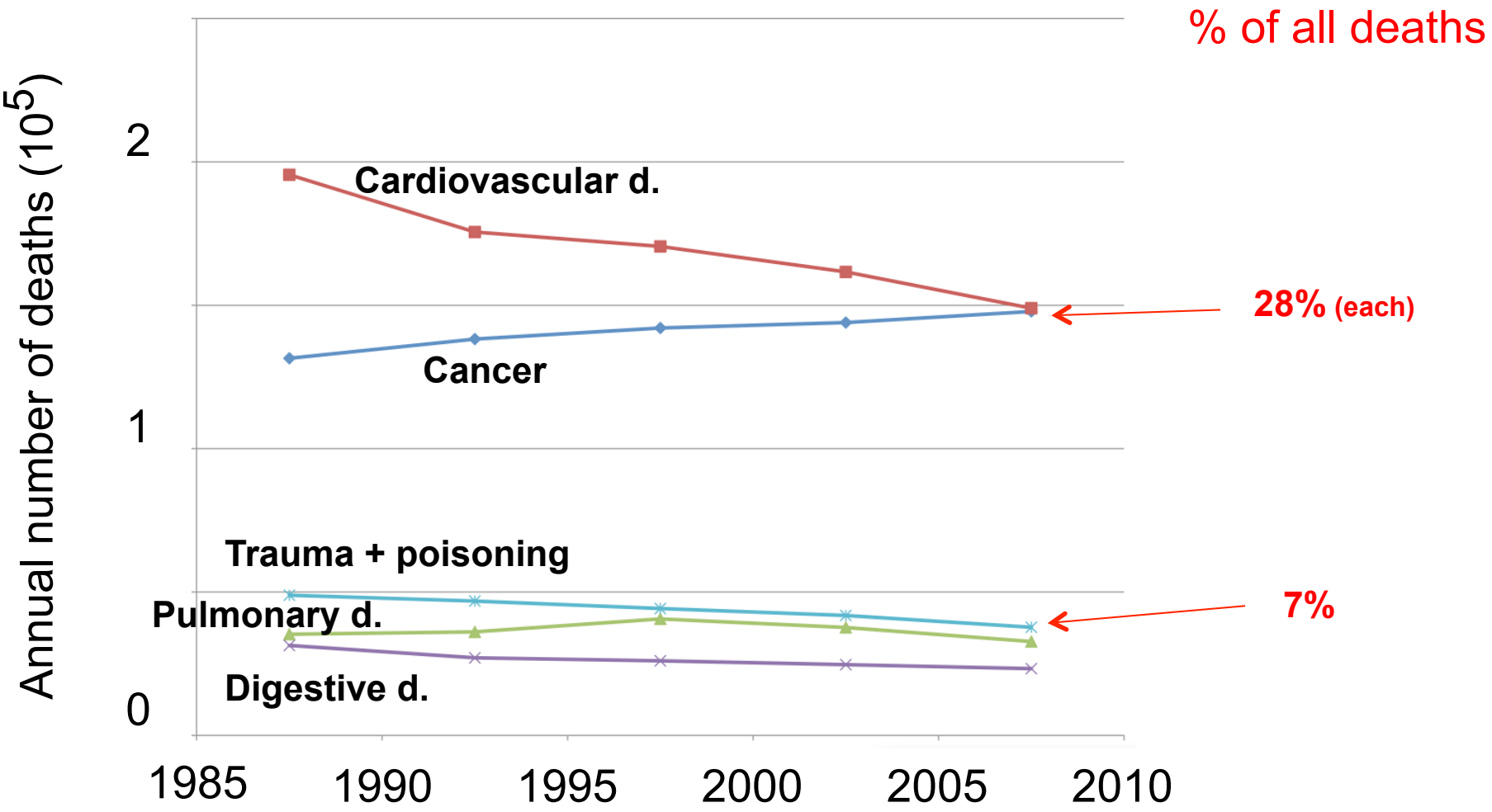
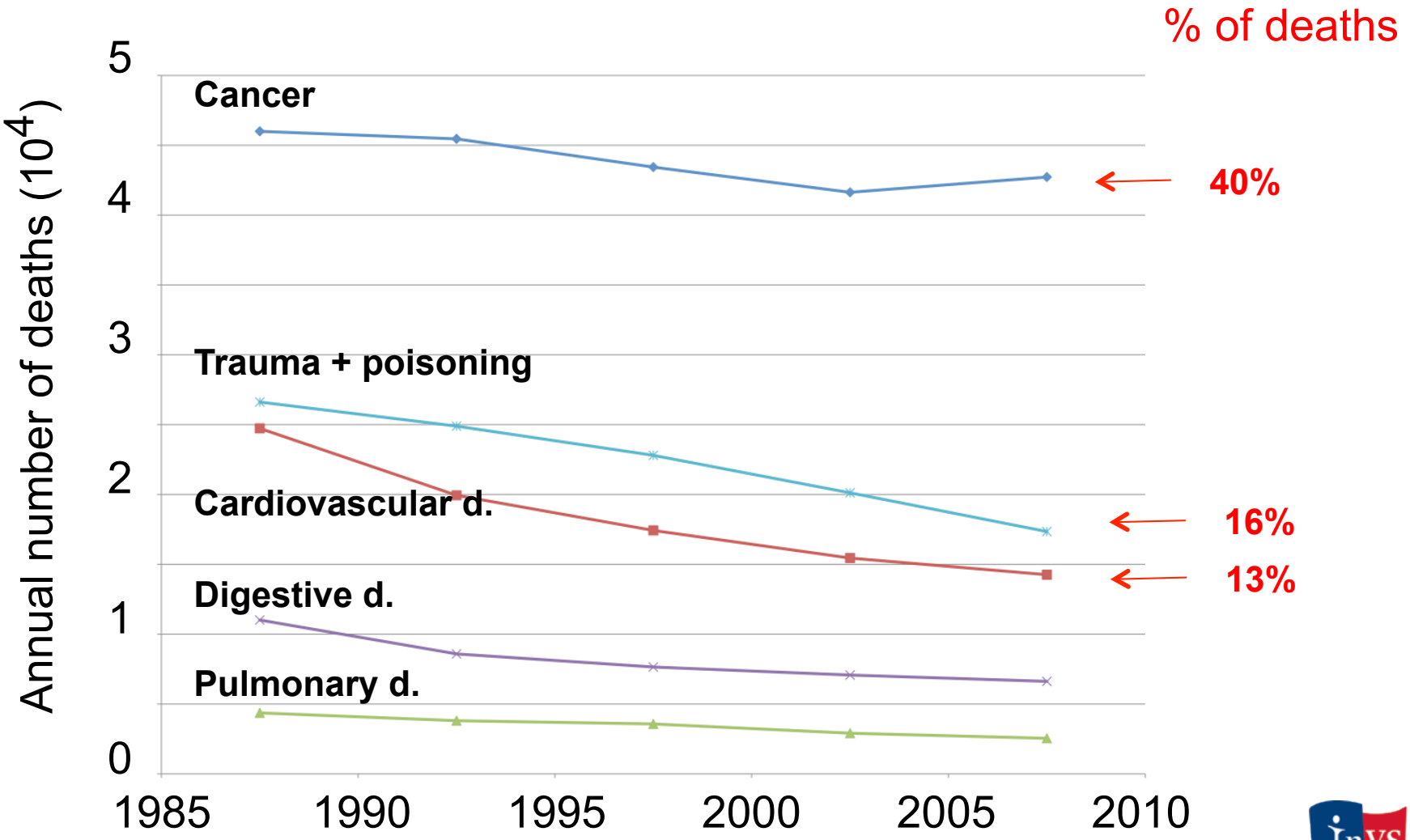
Main causes of death in France



INSTITUT DE VEILLE SANITAIRE

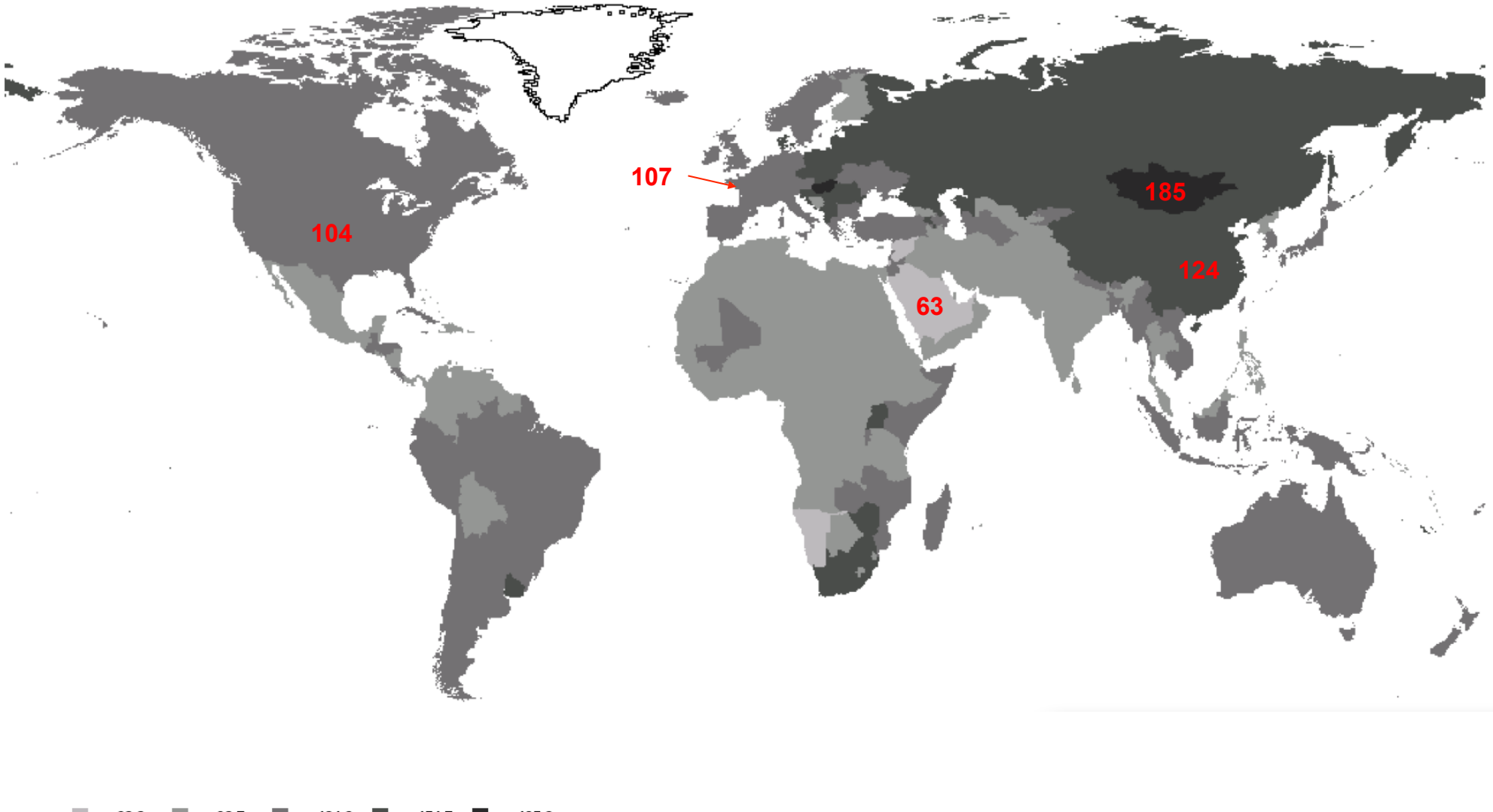
Main causes of “premature” death in France



Premature < 65 years

INSTITUT DE VEILLE SANITAIRE

Estimated age standardized mortality rate from cancer per 100 000



Globoscan - IARC

Causes of Cancer
Observational approaches

Environmental factors

Secular variation
in fixed populations

Migration studies

Analytical epidemiology

Intervention studies

Genetic factors

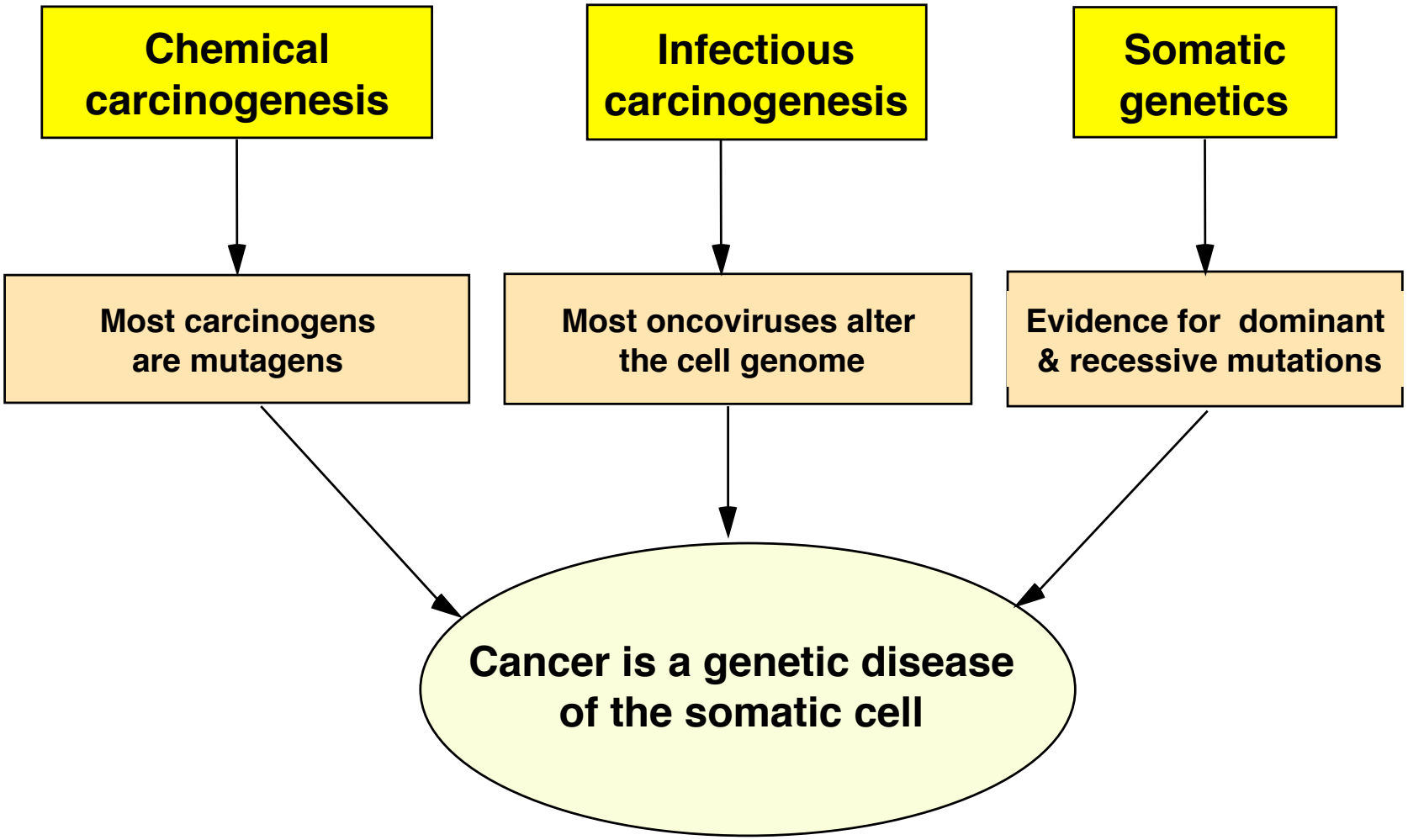
Twin studies

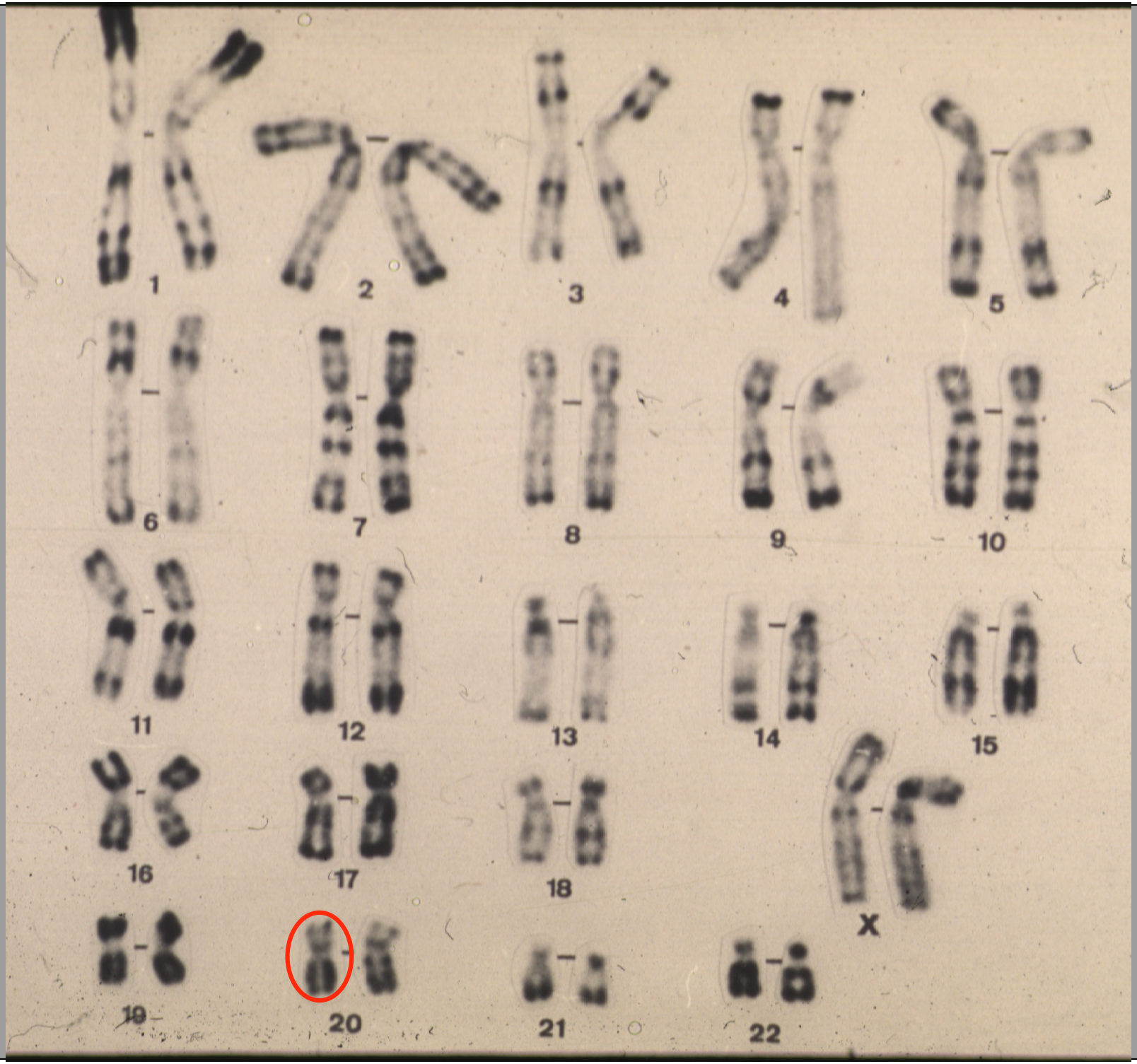
First degree relatives

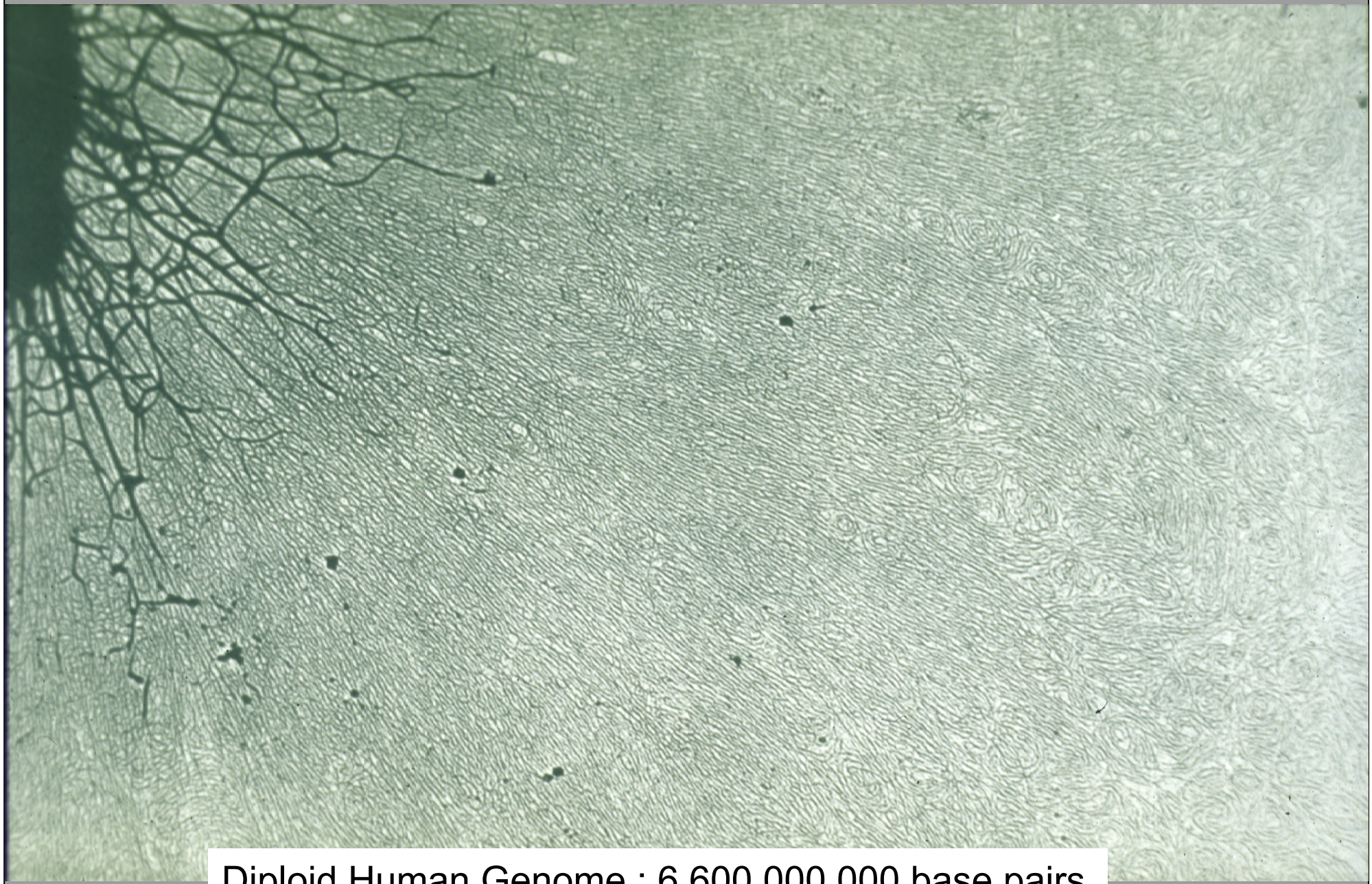
Cancer families




Causes of Cancer
Experimental approaches




1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X


Diploid Human Genome : 6 600 000 000 base pairs

Traditional strategies to identify candidate "cancer" genes

Retrovirology

- Transduced cellular gene
- 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

- Pathway
- Gene familie

Many routine tests for the management of cancer patients search for germline or somatic mutations

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

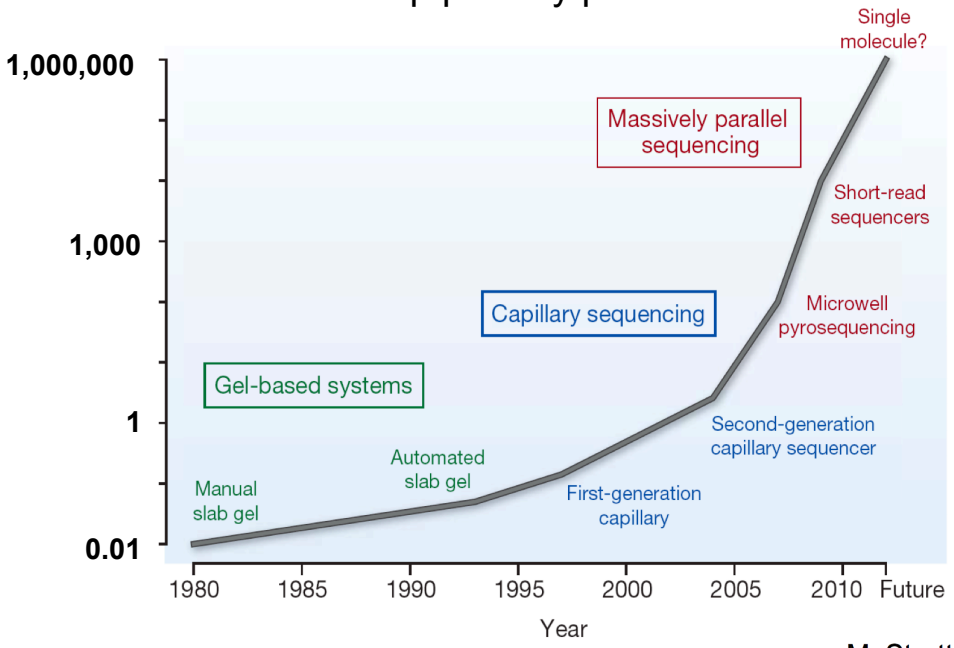
Response to treatment

- 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

The irruption of Next Generation Sequencing

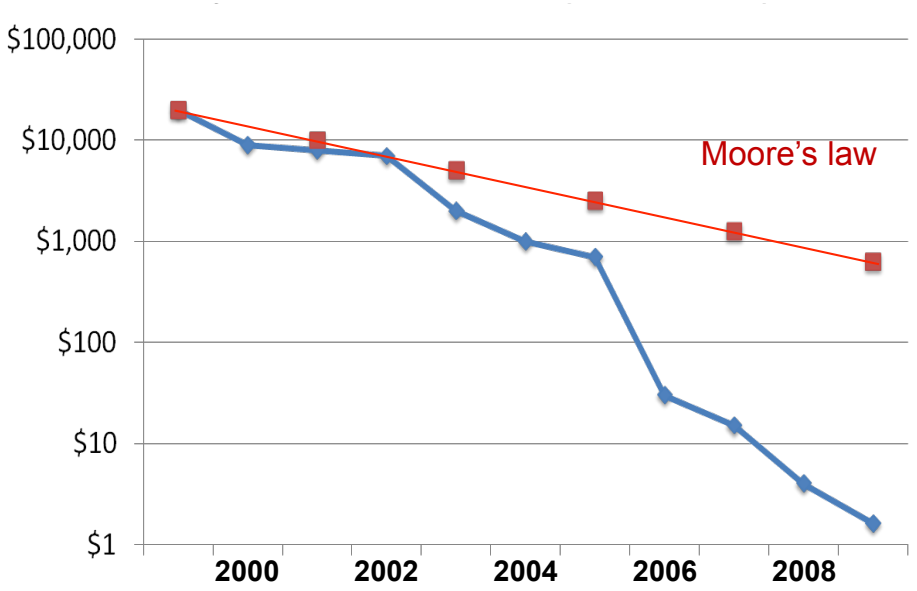
Increasing throughput

Million bp per day per machine



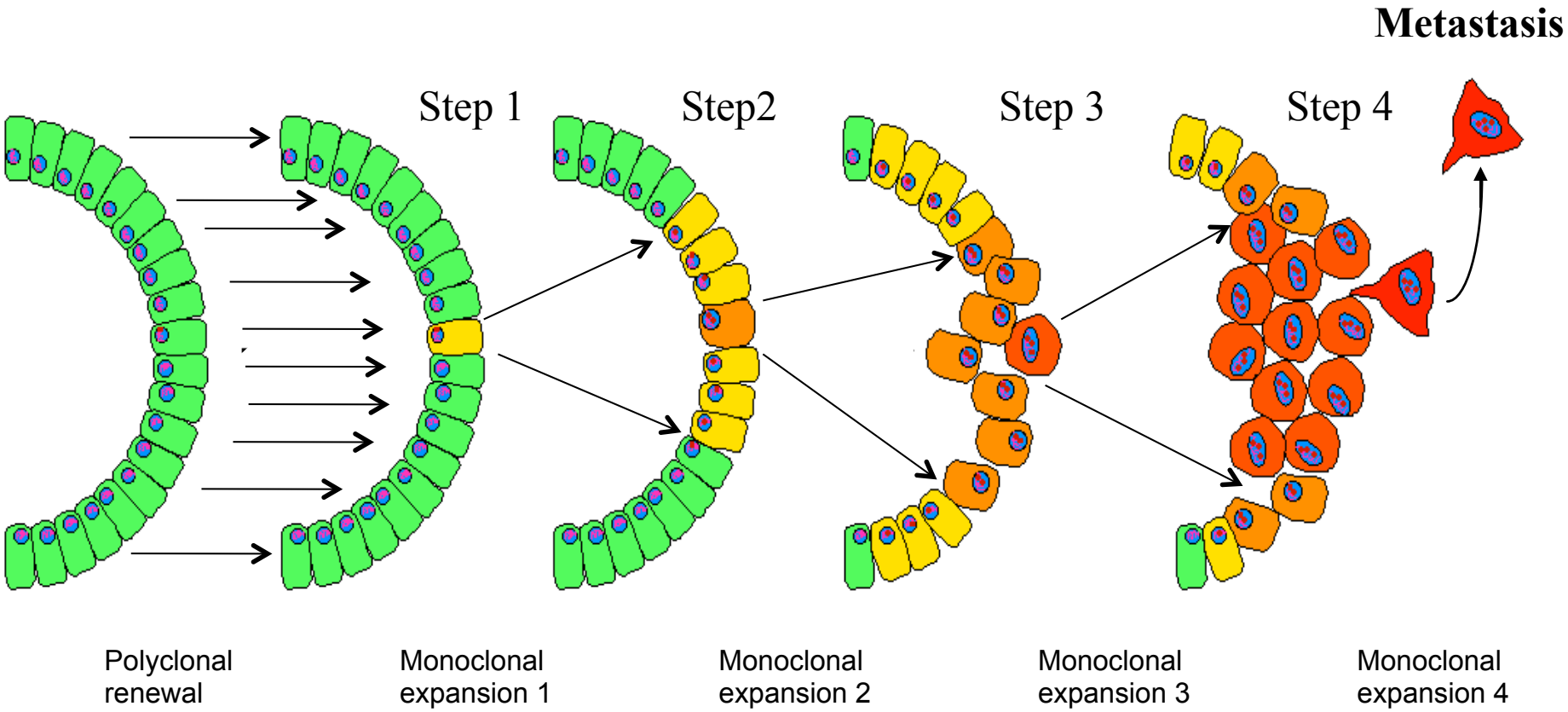
Decreasing cost

Cost per million bp sequenced

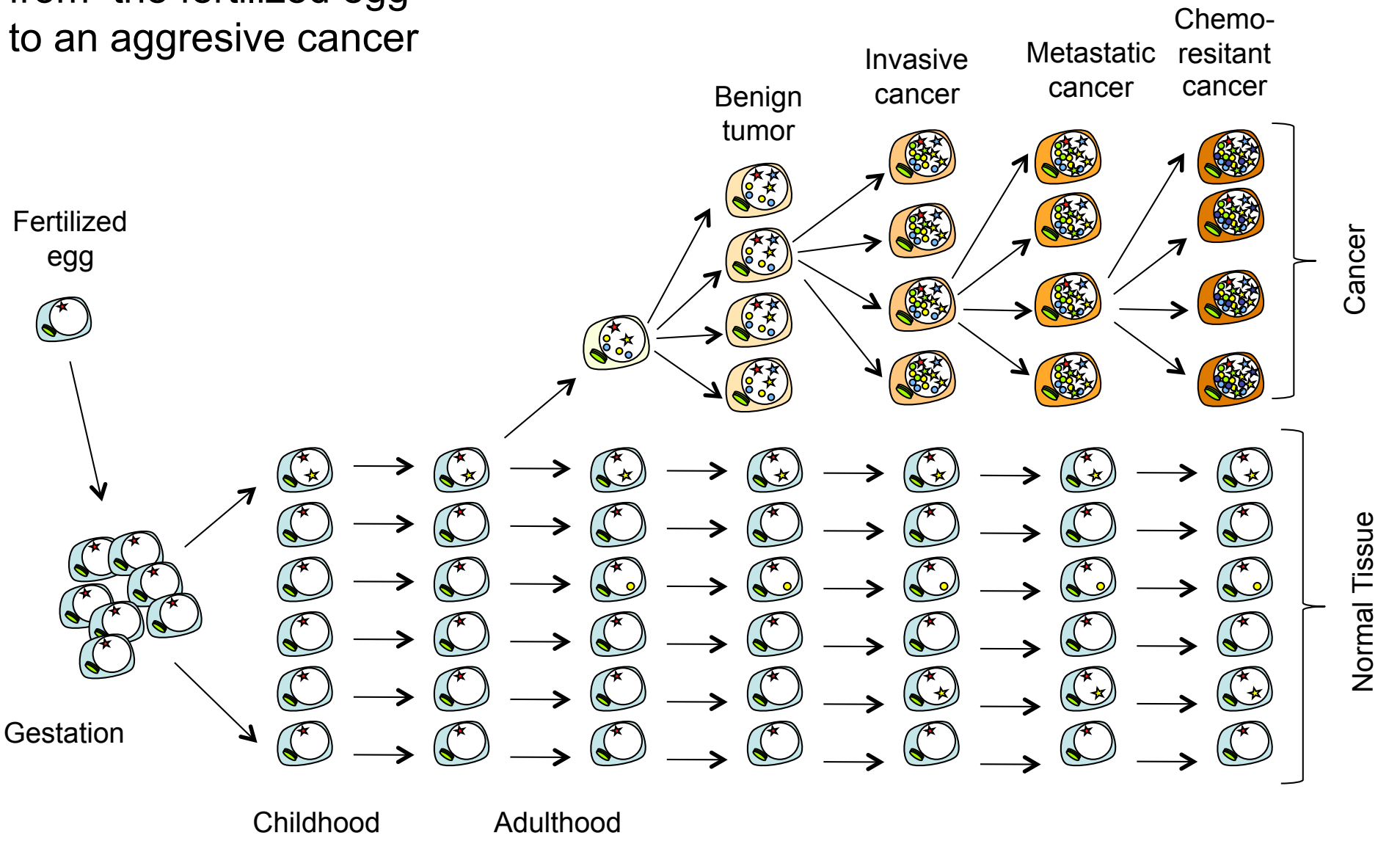


M. Stratton

Carcinogenesis : a multistep process

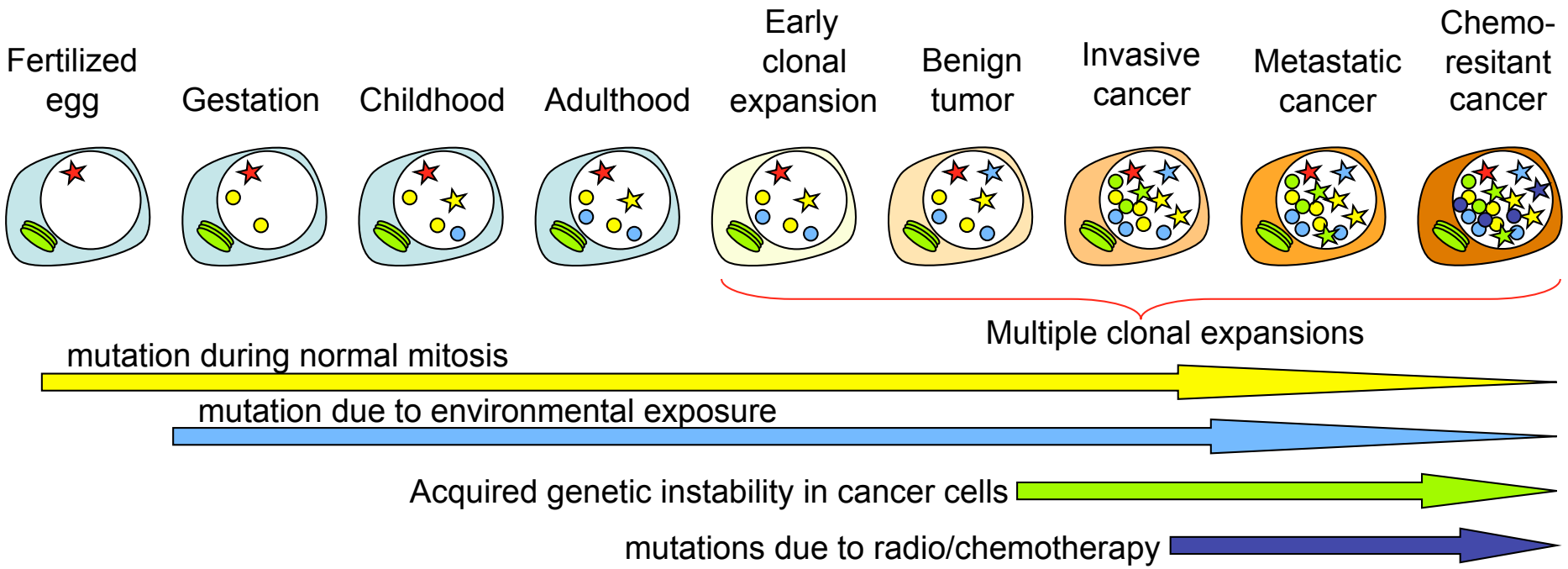


Somatic genetic events from the fertilized egg to an aggresive cancer



The "average" genome of blood cells provides a fair representation of the fertilized egg genome

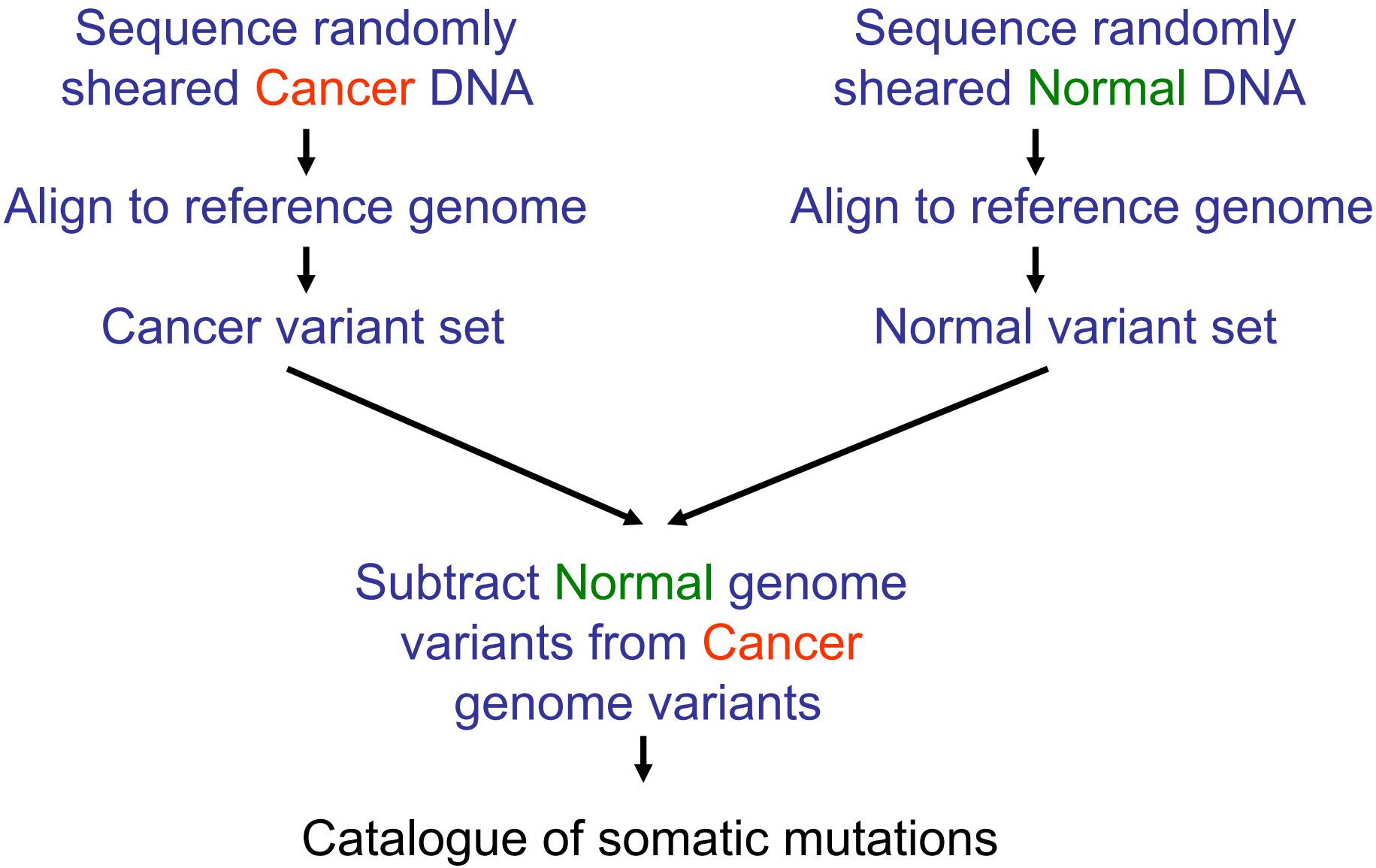
Somatic genetic events
from the fertilized egg to an aggresive cancer



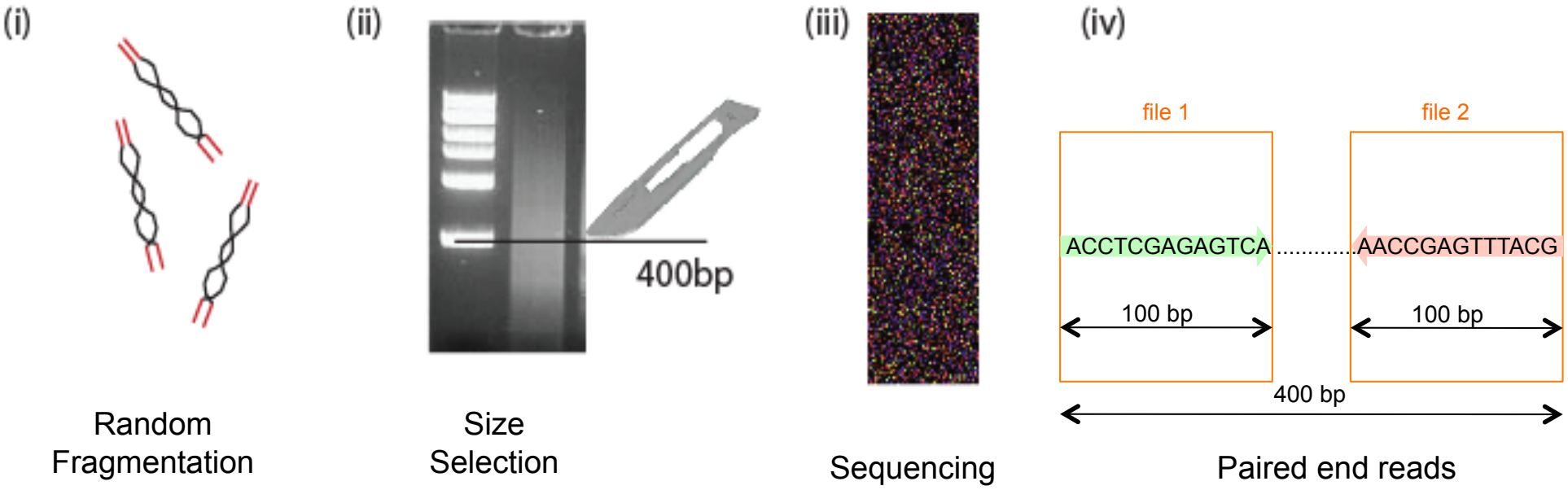
Passenger mutations
Driver mutations !

Adapted from M. Stratton

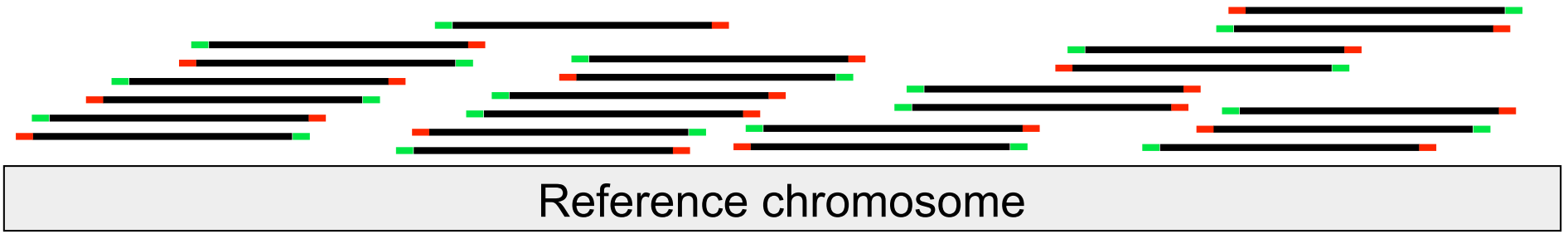
Overall strategy for detection of somatic mutations



Paired-end sequencing



Map sequences back to reference genome



Present standard : reliable reconstructionr equires average sequence coverage > 30

Total sequence to be acquired : 90 000 000 000 base pairs
(Illumina : 900 000 000 sequence reads)

Acquisition and analysis of whole genome sequences for a matched pair of normal/tumor DNAs

HiSeq sequencer

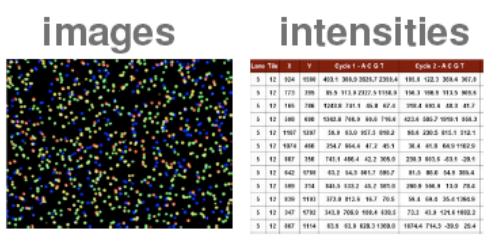
9 days

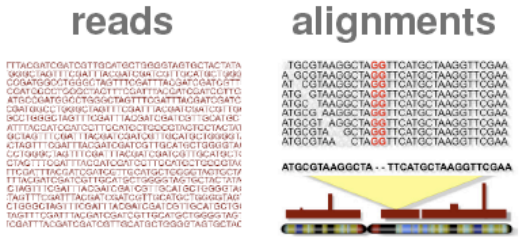
Data quality control

4 days

Identification of somatic mutations

Several weeks





Genome Reconstruction & comparison

32 files

400 GiBytes

256 Cores ,

RAM : 768 GBytes

Multiple algorithms in development

Requires independent validation

Mostly standardized

Mostly automatic

Partly manual

Types of genomic alteration acquired by cancer cells during tumorigenesis

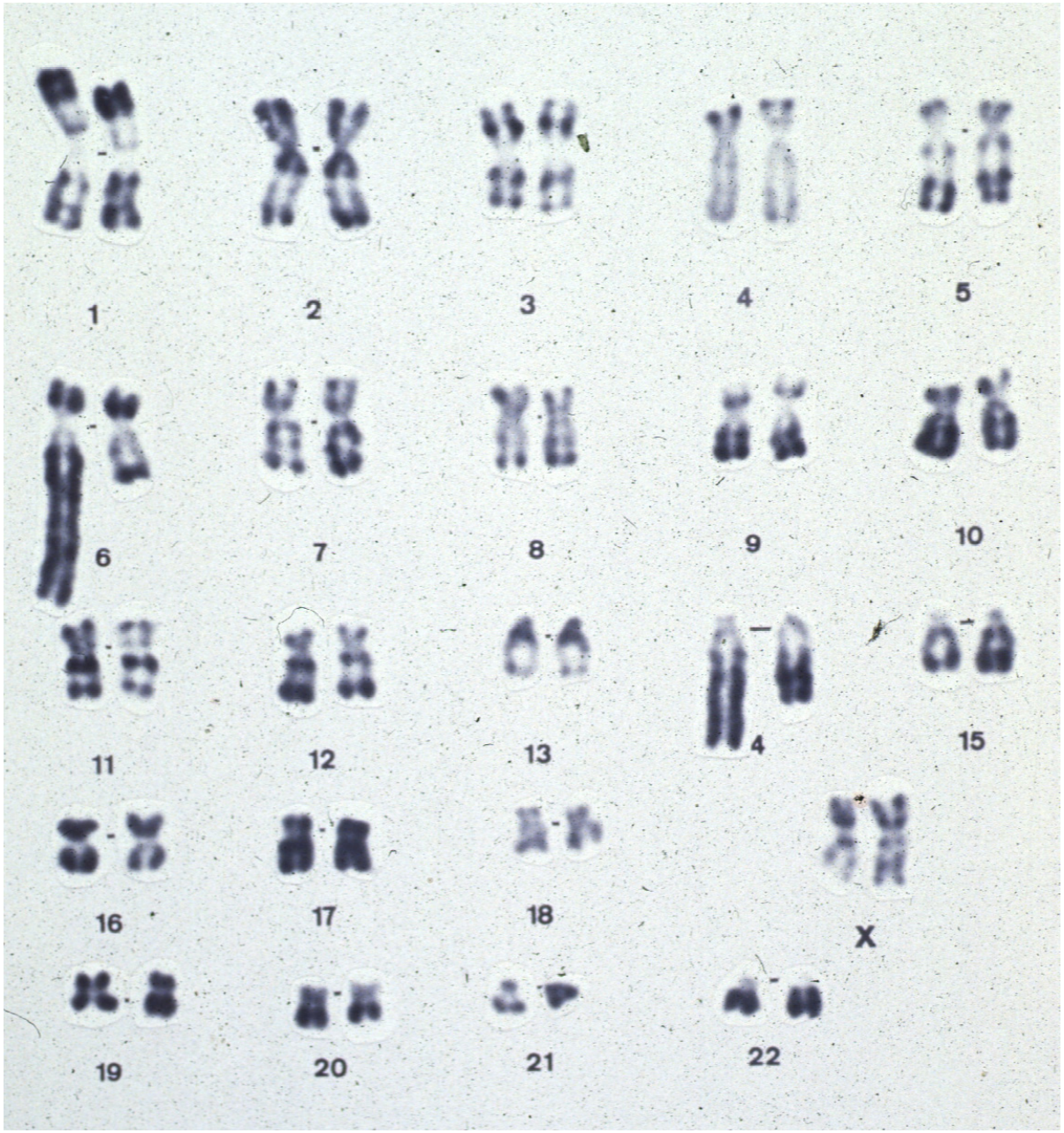
Structure

- Gene amplification
- Interchromosomal rearrangements
- Intrachromosomal rearrangements
- Point mutations

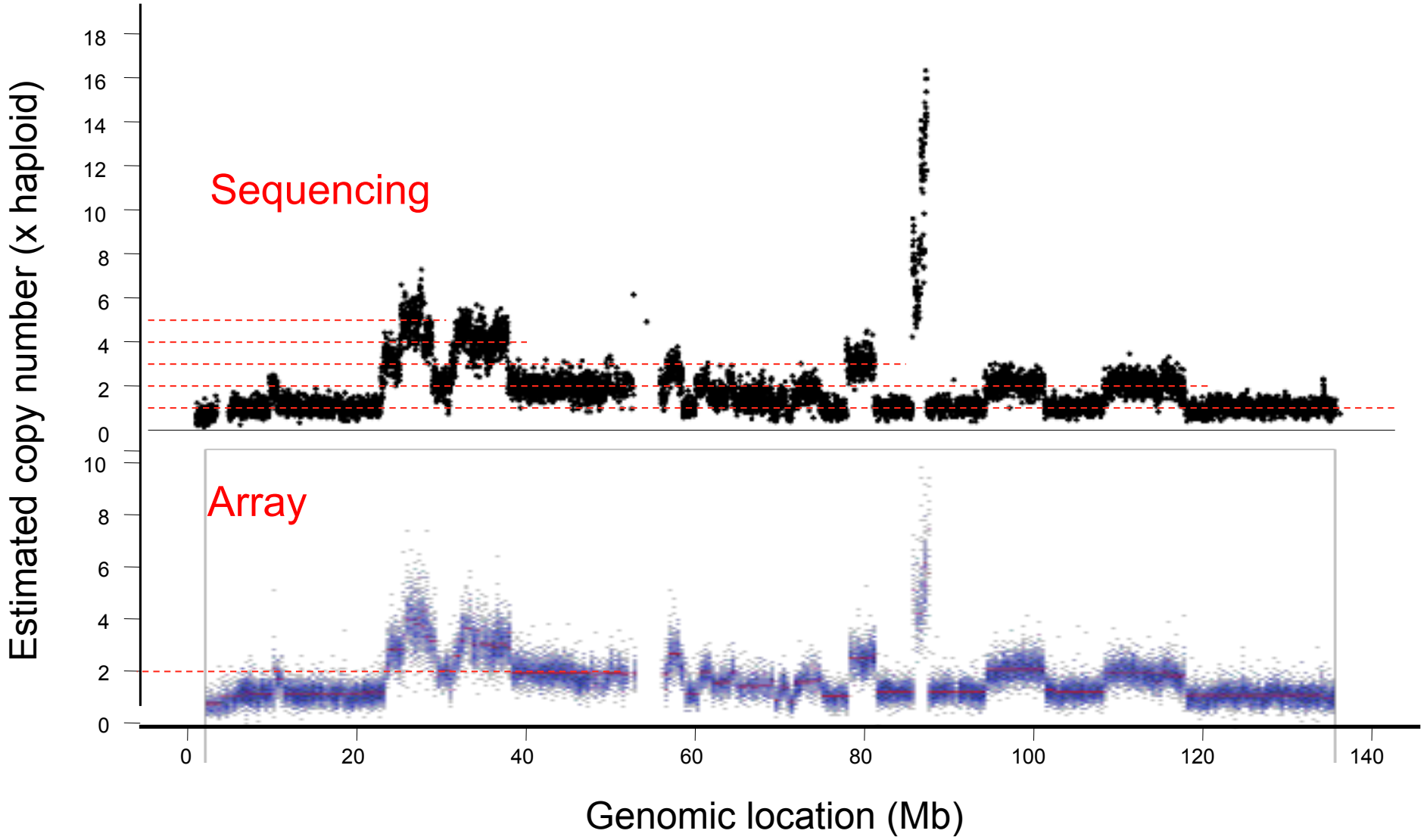
Function

- Driver mutations ⟹ “driving” the clonal expansions
- Passenger mutations ⟹ “hitchhiking” the clonal expansions

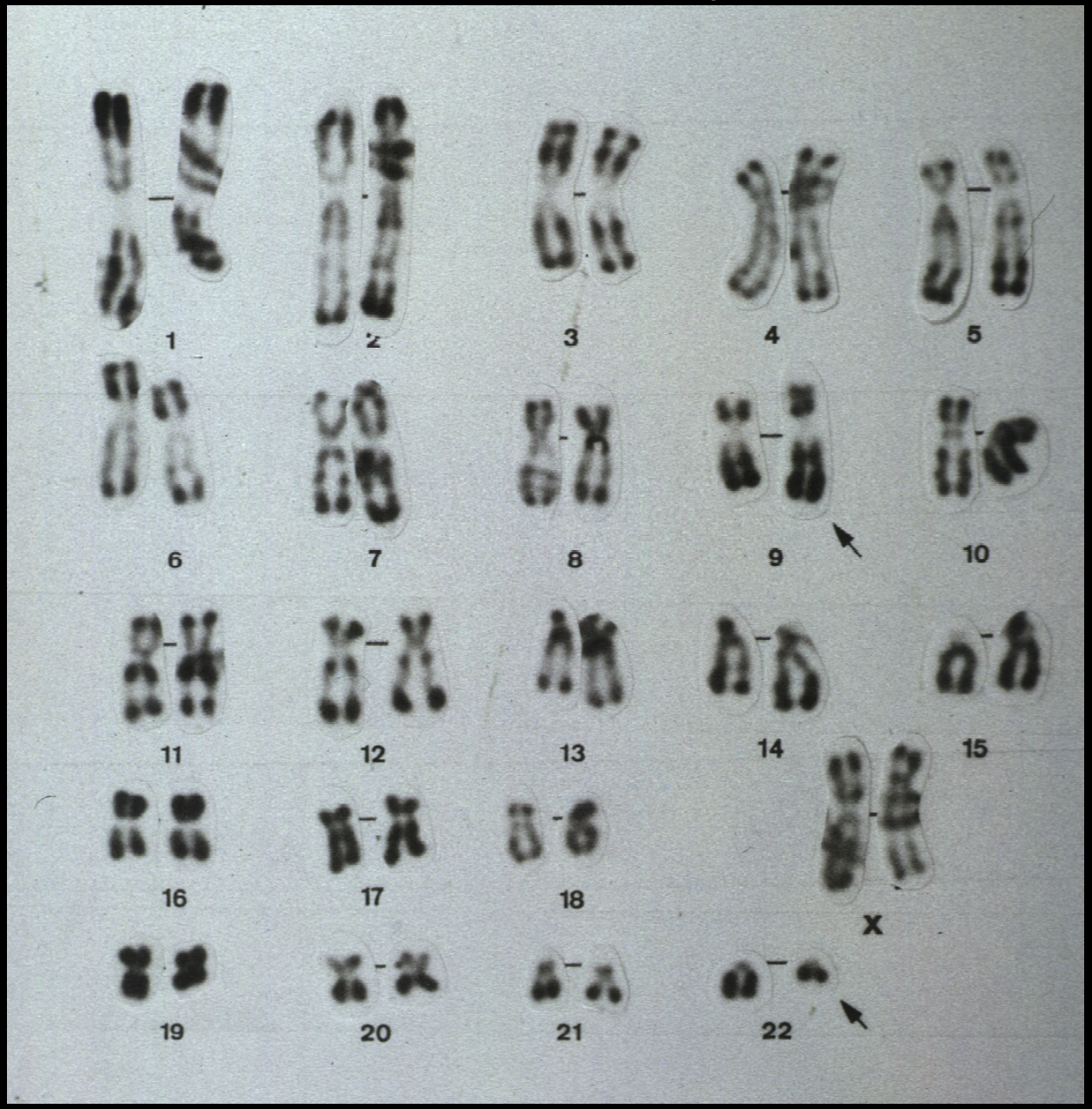
Neuroblastoma

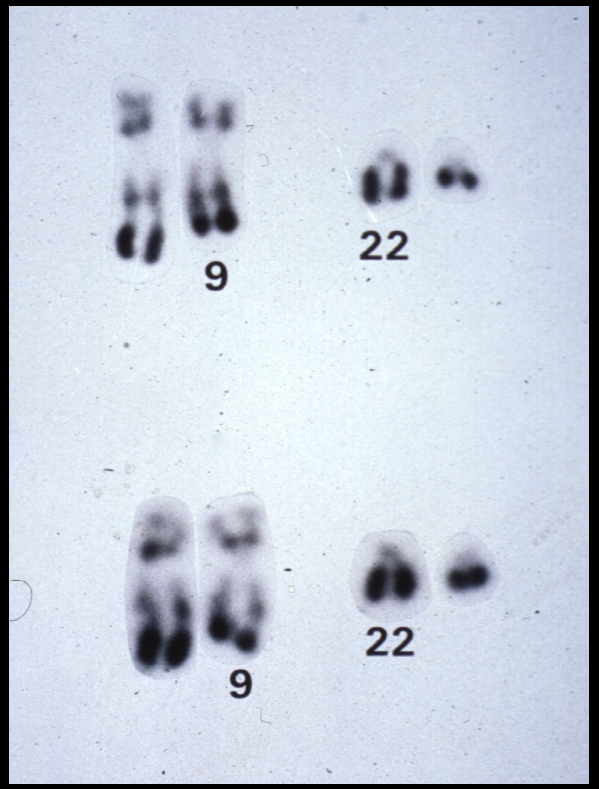


Comparative evaluation of copy number variation by sequencing and DNA arrays

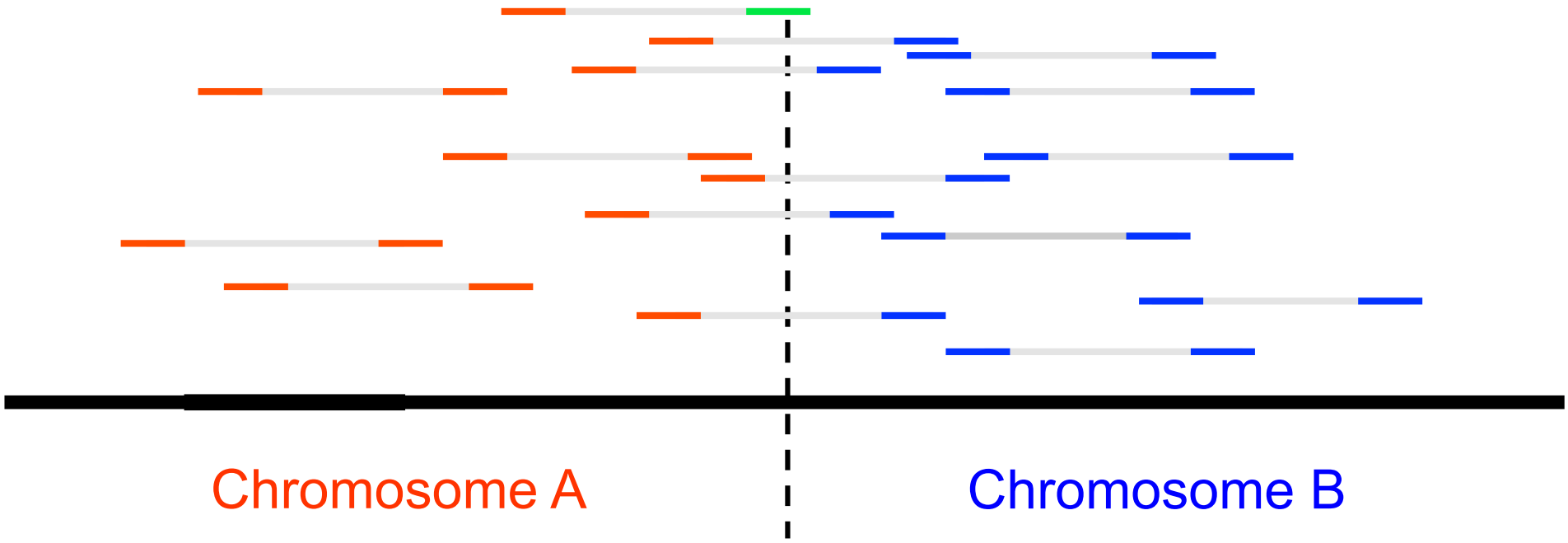


Chronic myeloid Leukaemia

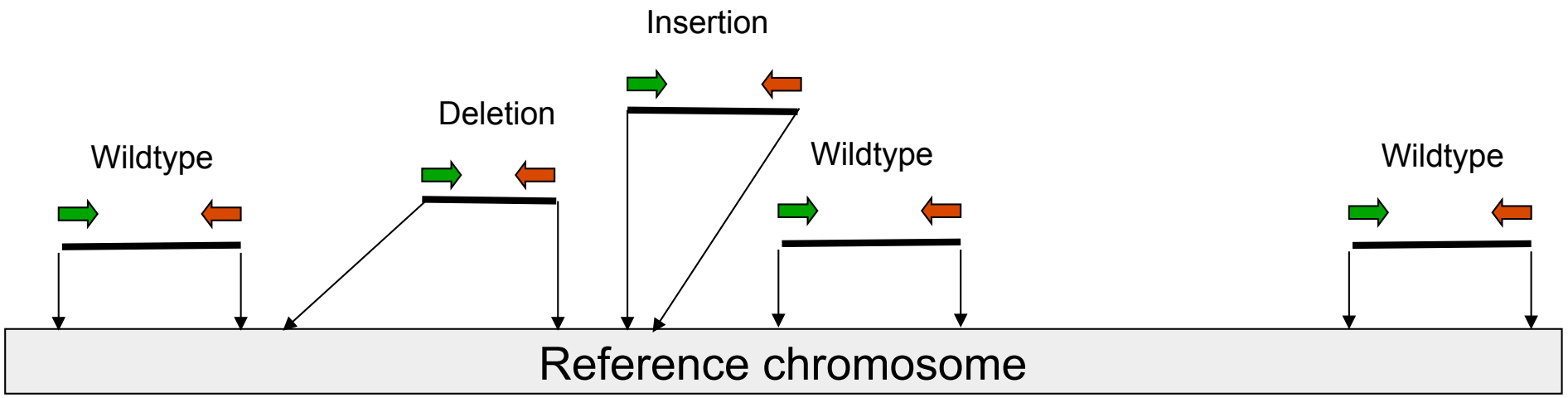




Search for interchromosomal rearrangements (translocations)

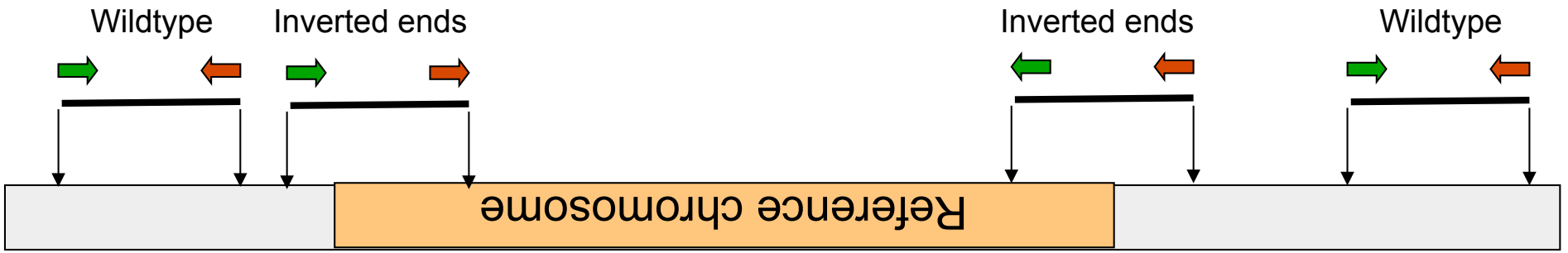


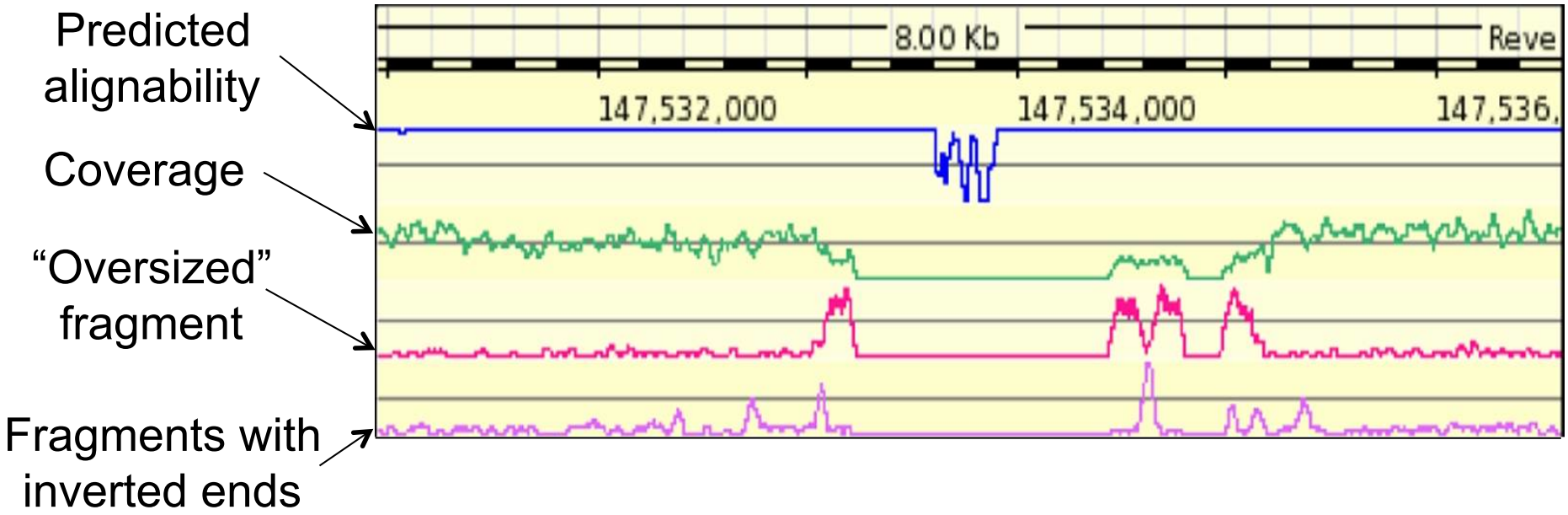
Search for intrachromosomal rearrangements



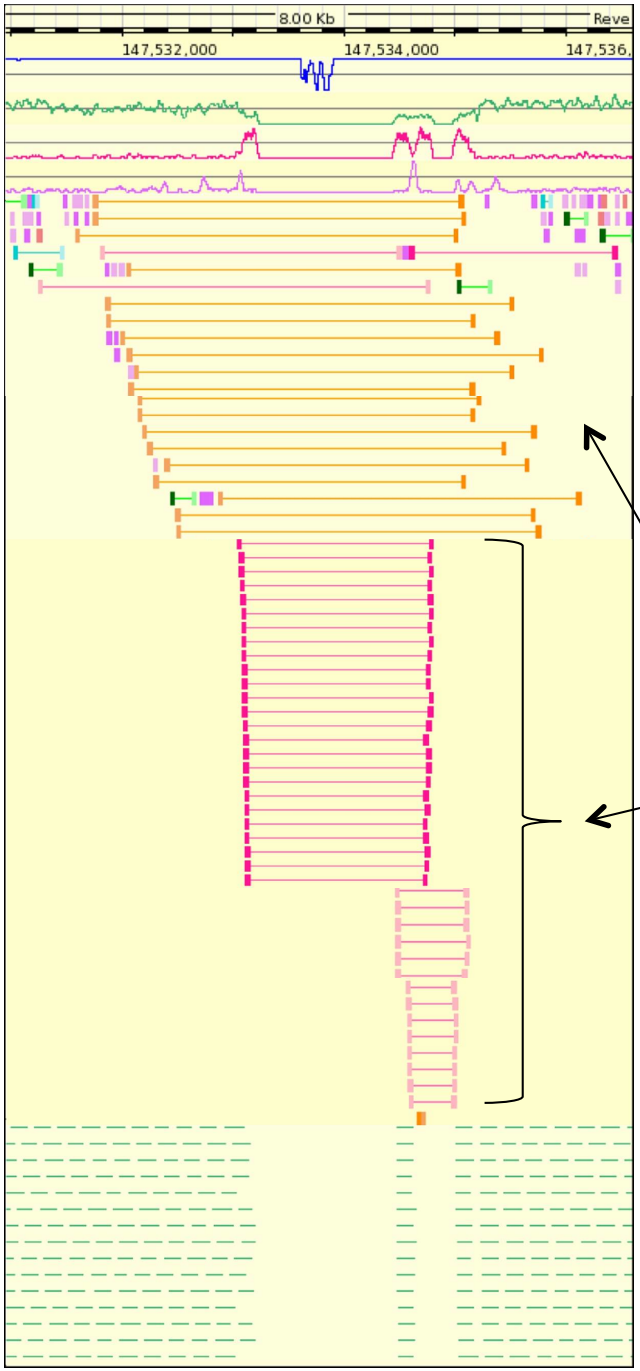
size fractionated 400 bp random fragments of cancer genome

Chromosome Inversion





Bentley Nature 2009



Homozygous Complex Rearrangement

Deletion
1206 pb

Inversion
369 pb

Deletion
164 pb

reference genome

sequenced genome

Normal ends

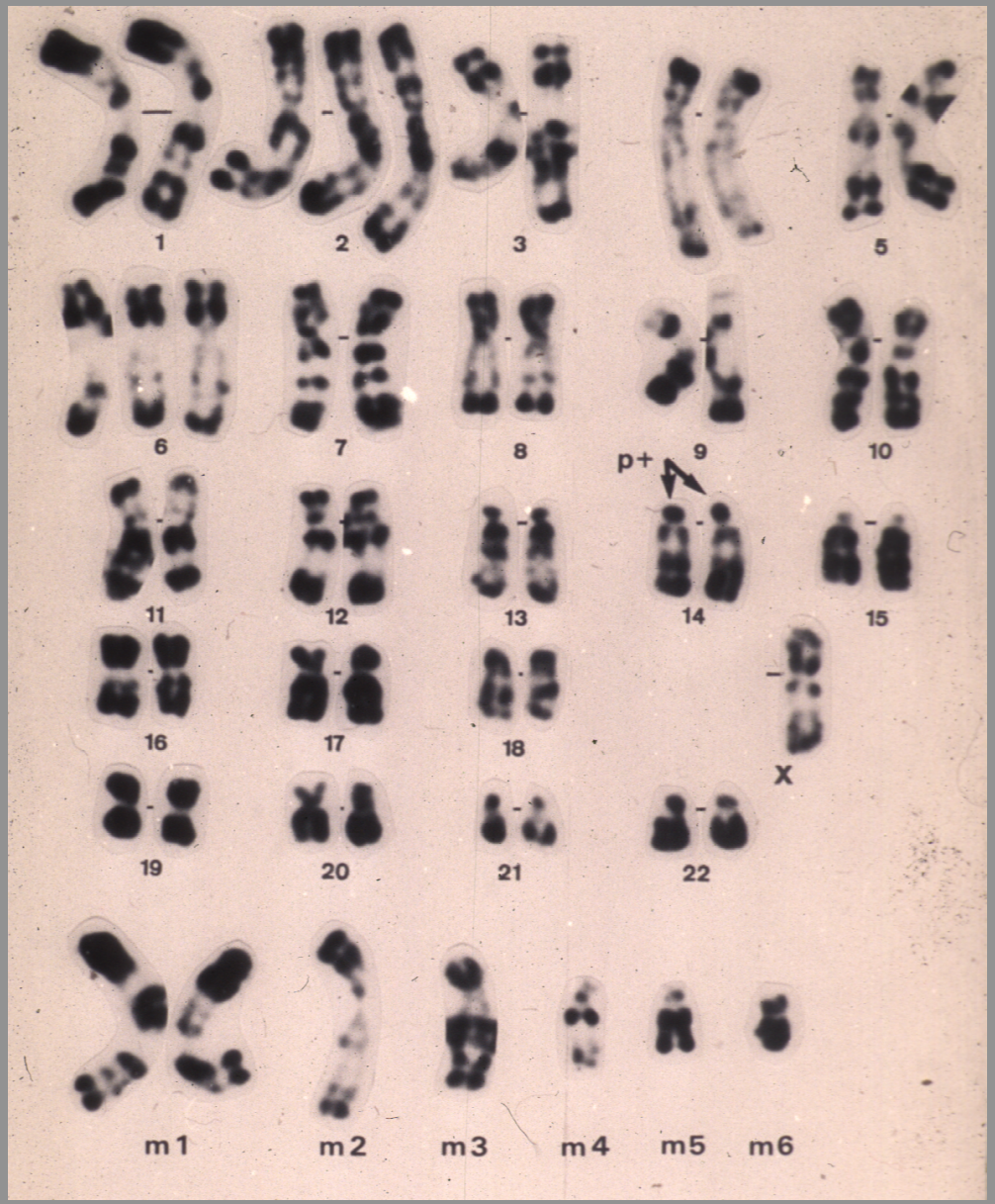
Inverted ends

Too long

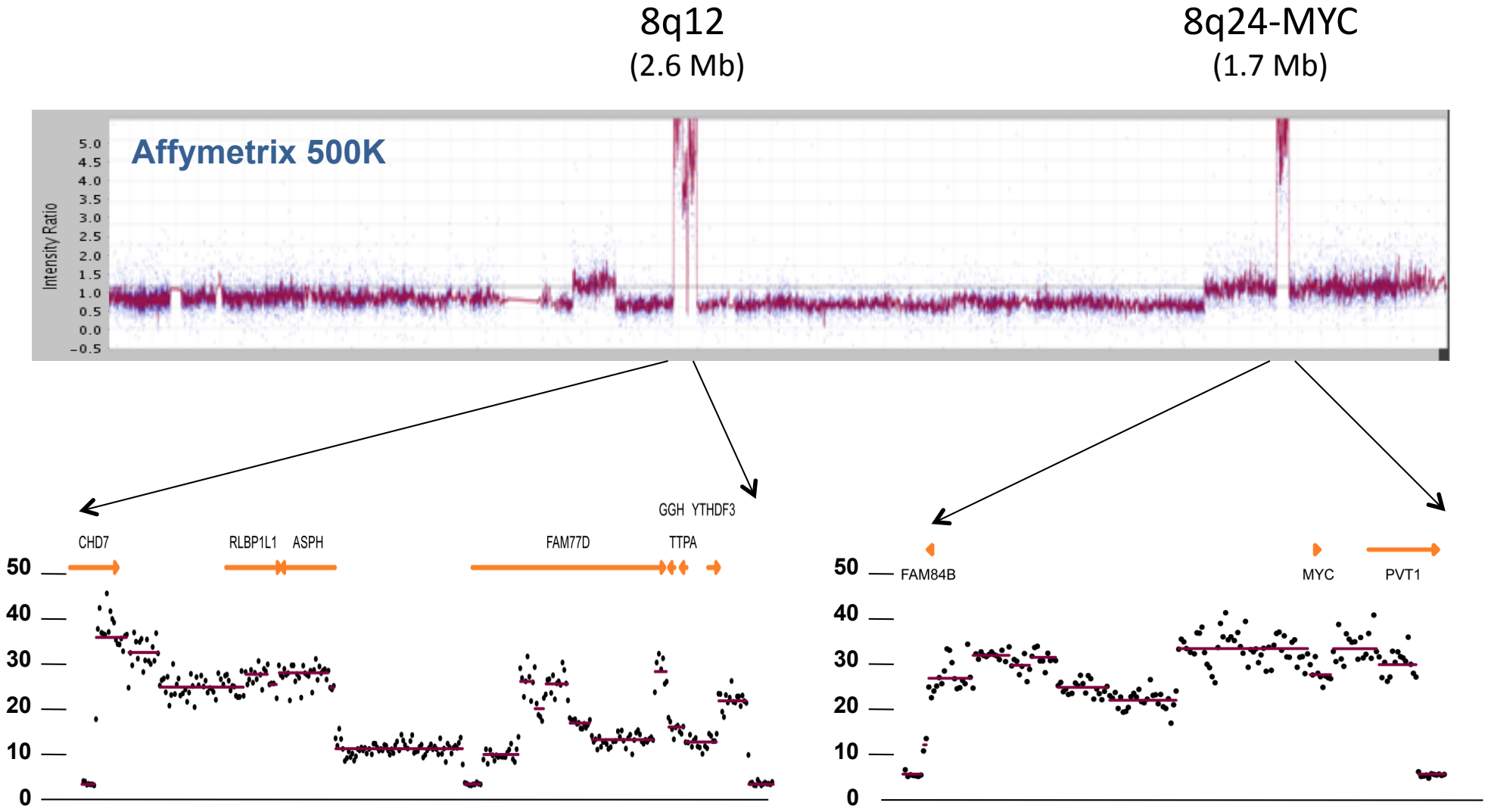
Expected

Bentley Nature 2009

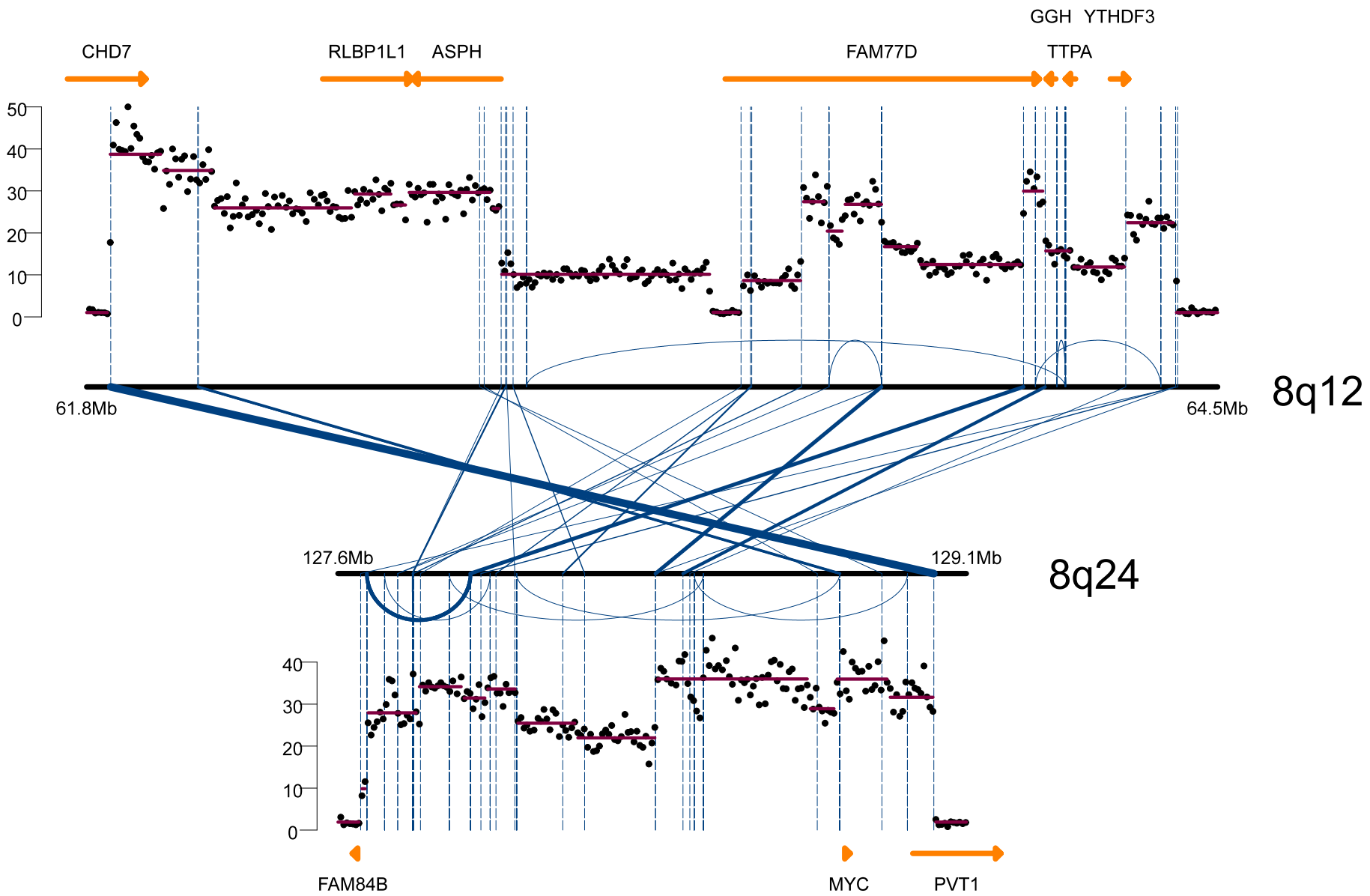
Complex karyotype



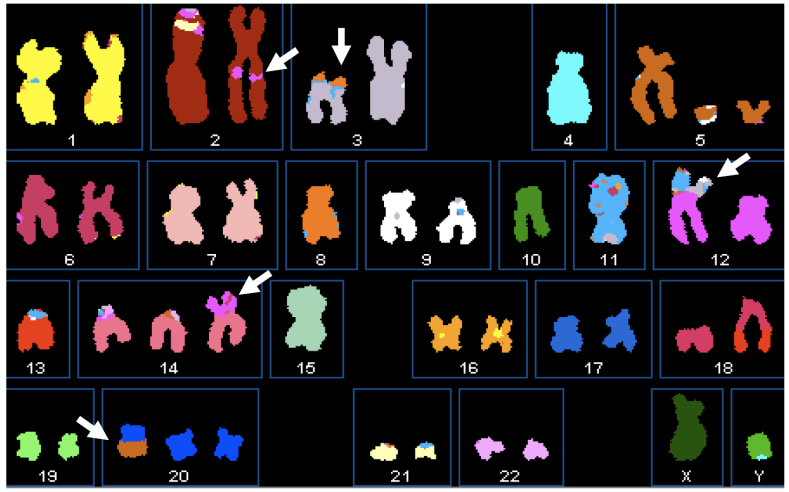
Single-end mapping of short reads



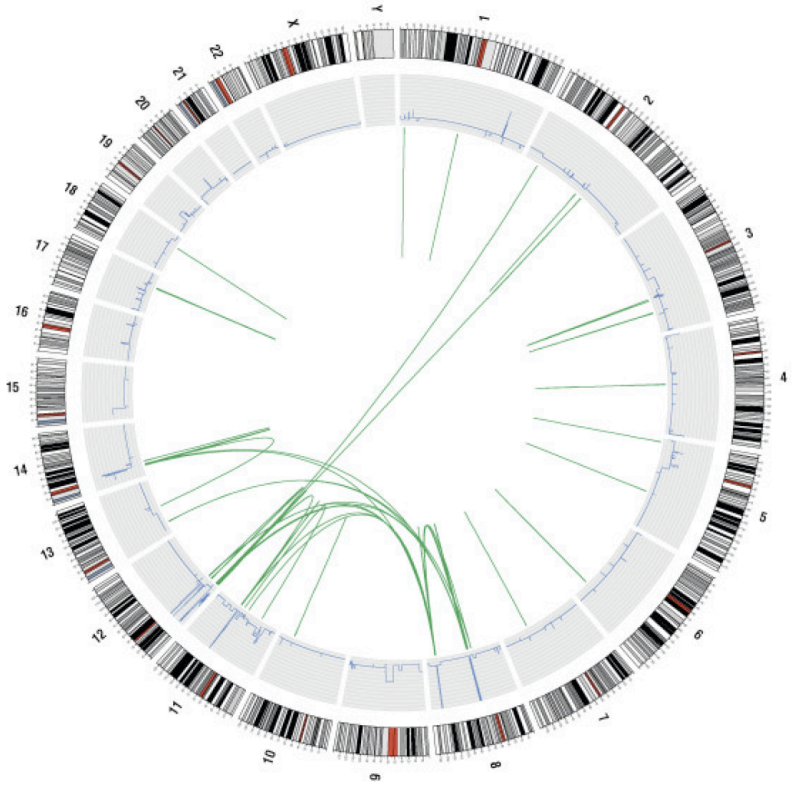
Paired-end mapping of short reads



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



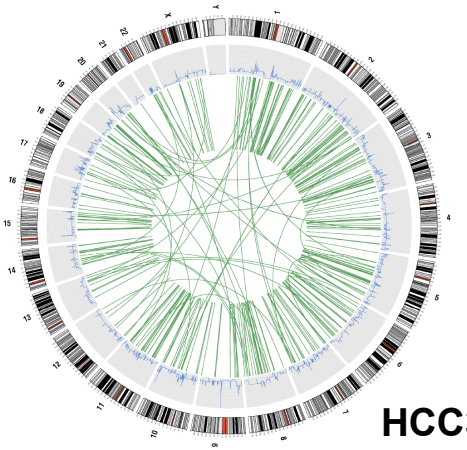
Spectral Karyotyping
~10 chromosome rearrangements
(all interchromosomal)



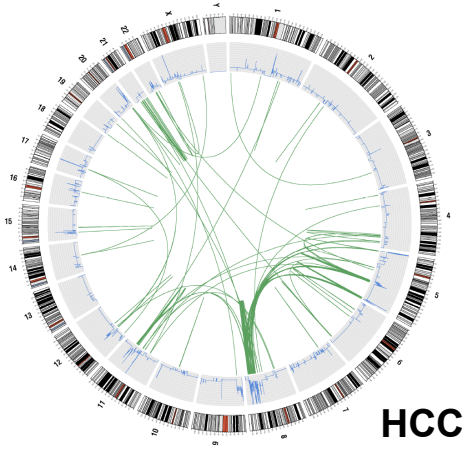
Paired end sequencing
88 chromosome rearrangements
(59 intrachromosomal, 22 interchromosomal)

Courtesy M. Stratton

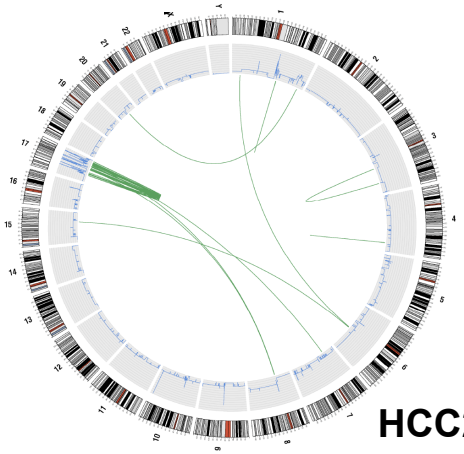
Chromosome rearrangements in breast cancer



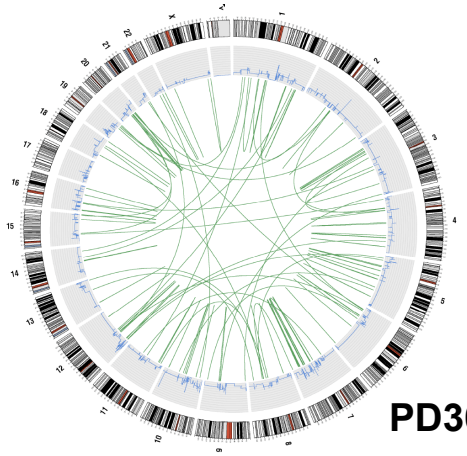
HCC38



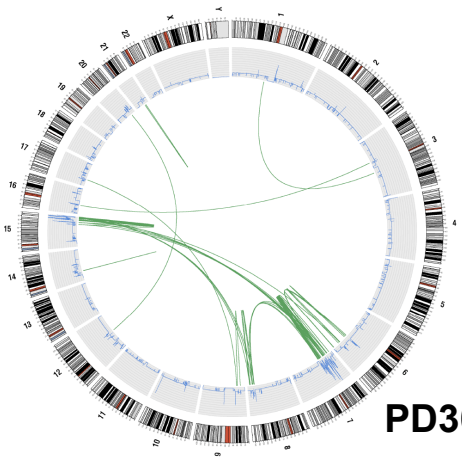
HCC1954



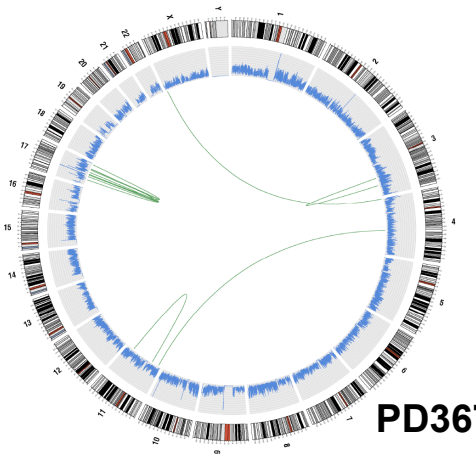
HCC2218



PD3665a



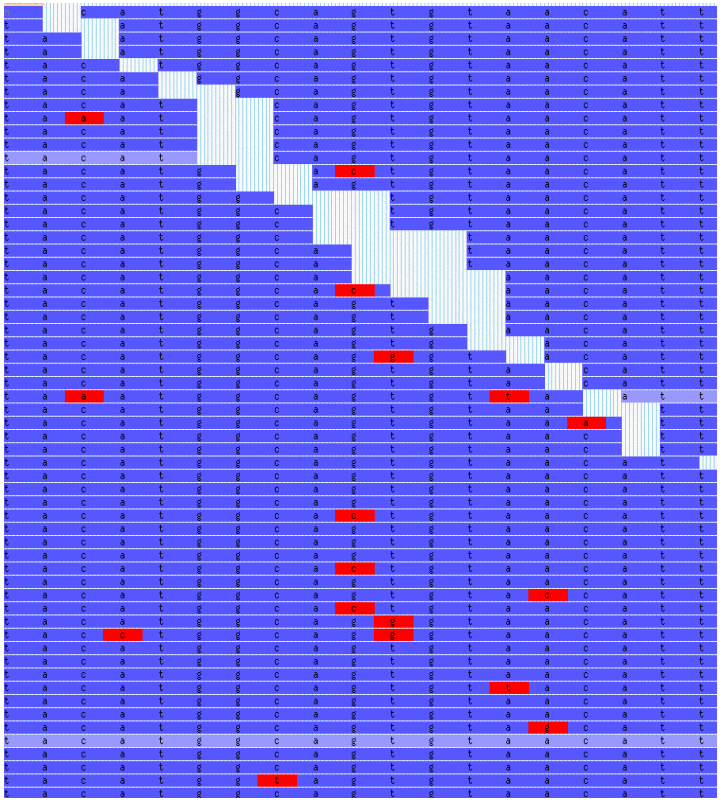
PD3668a

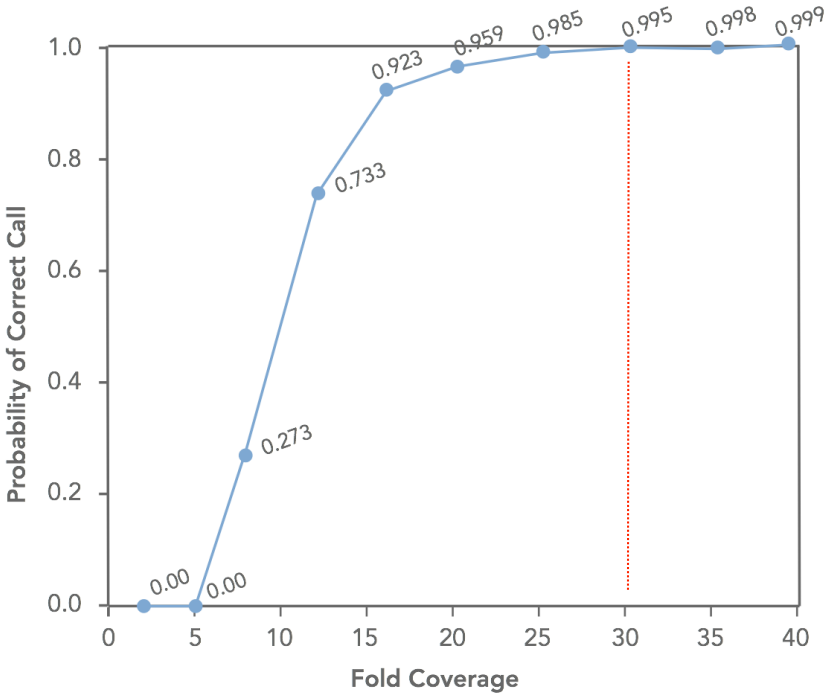


PD3671a

Courtesy M. Stratton

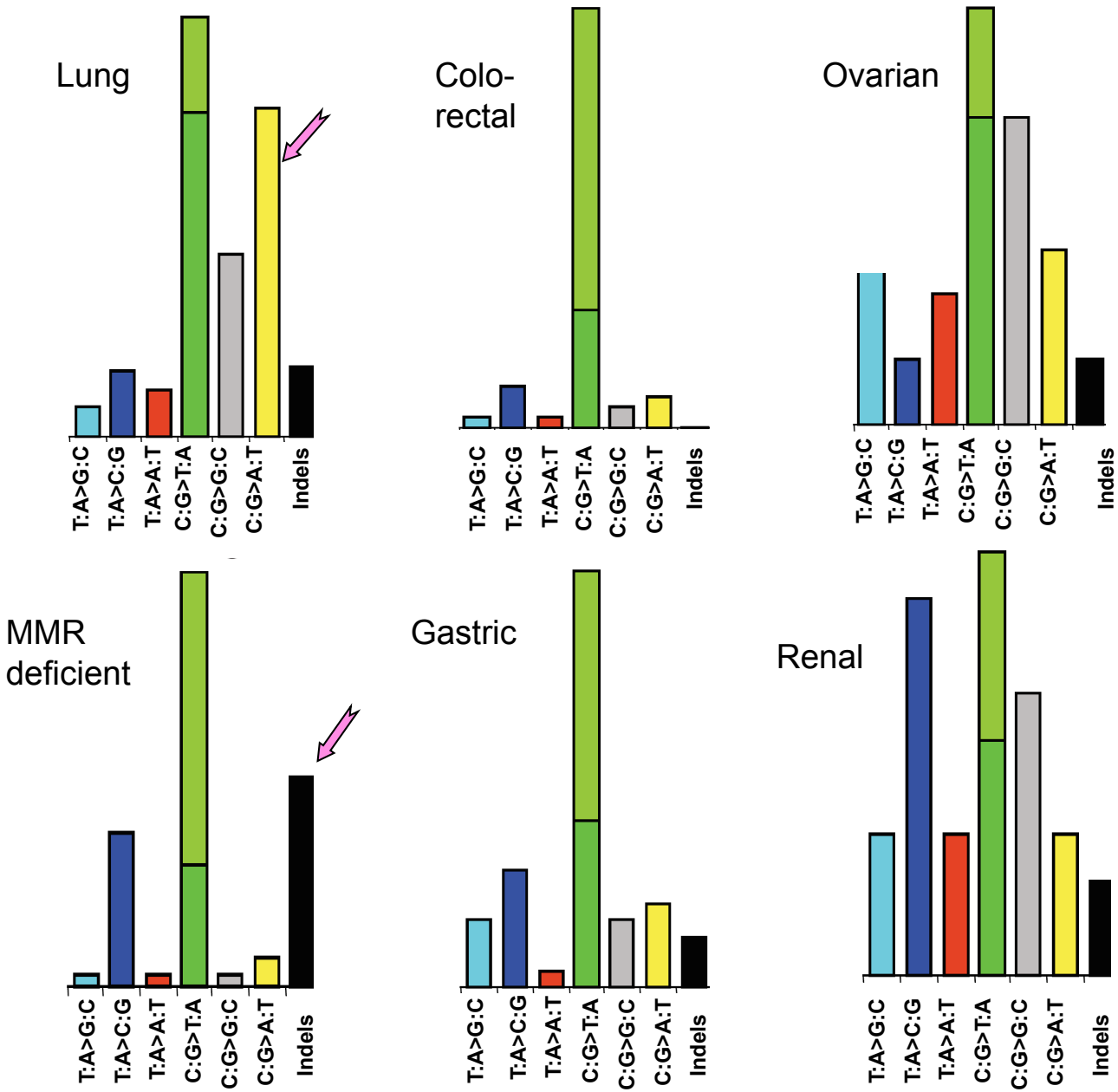
Detection of single nucleotide variants



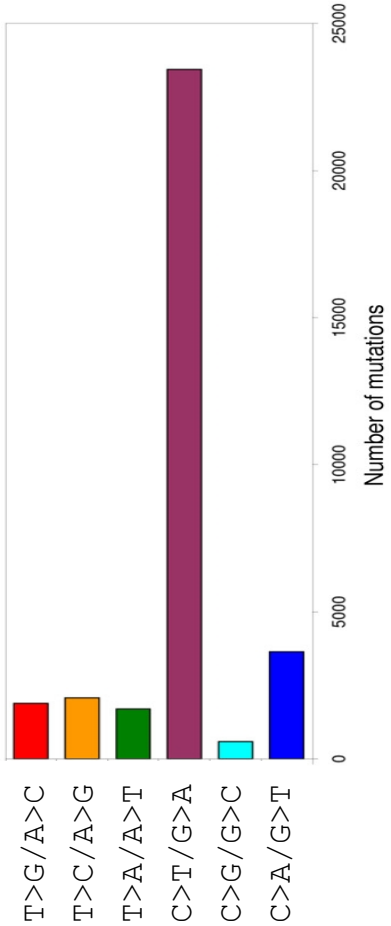


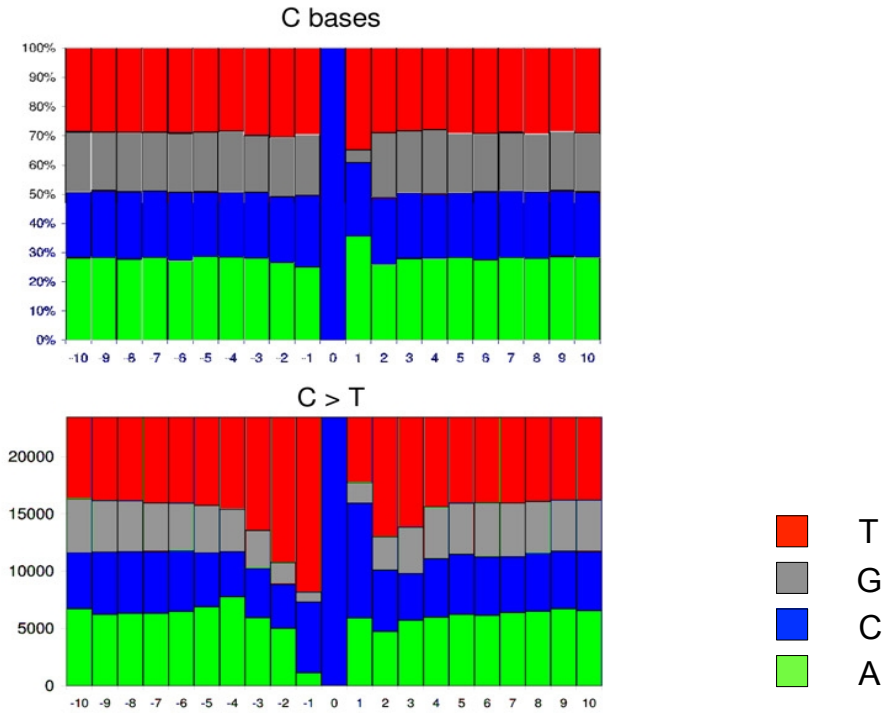
The number of somatic point mutations is highly variable among tumors, even within a single tumor type.

The spectrum of somatic mutations differs with tumor type

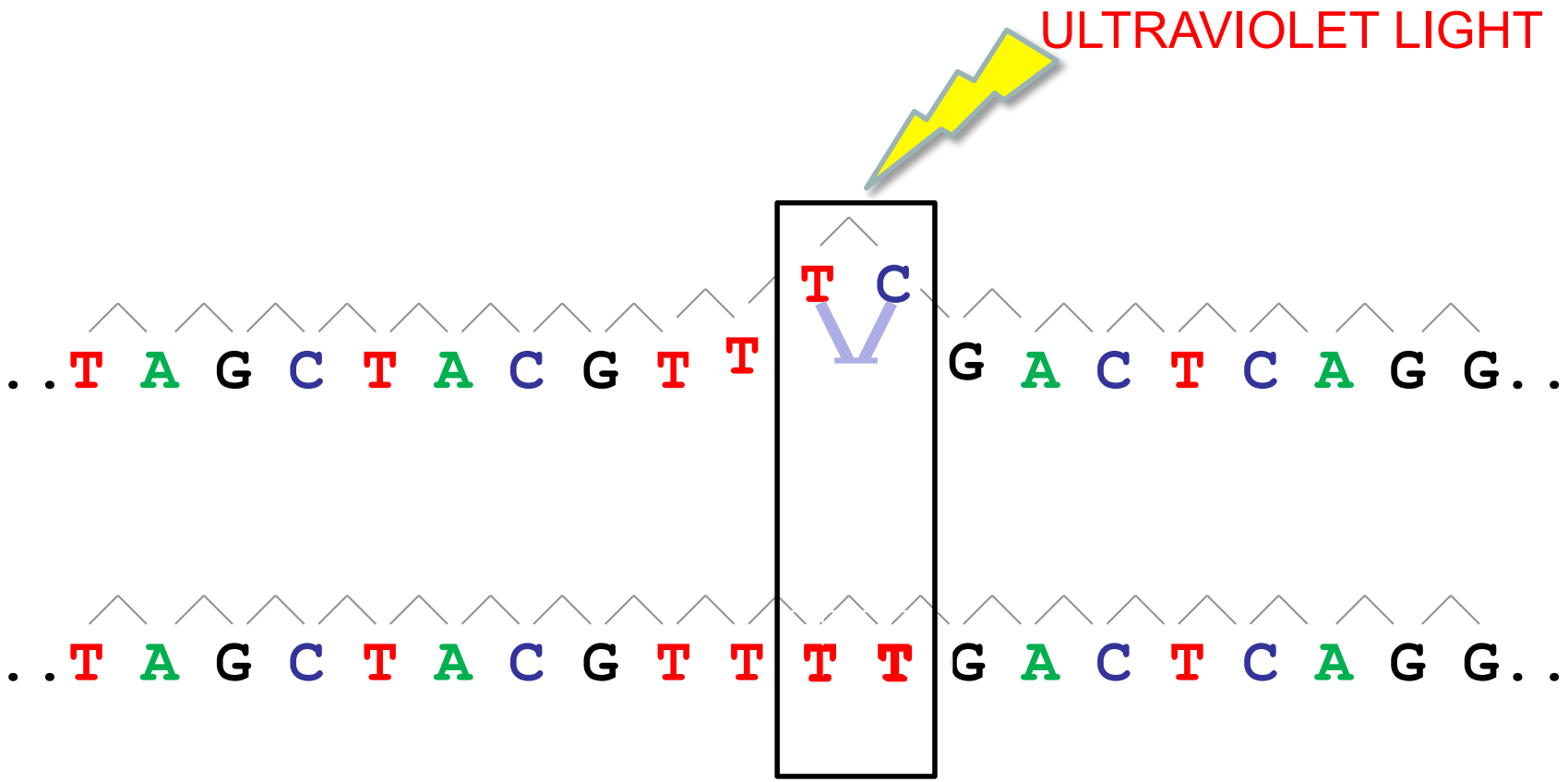


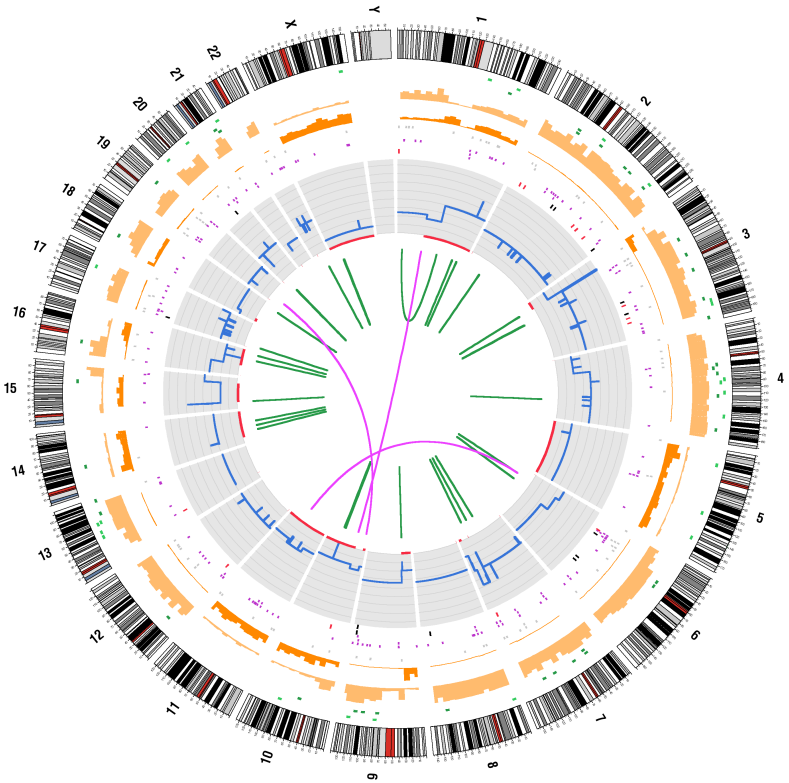
Sequence context of mutations in COLO-829 (melanoma)
33,345 somatic base substitutions

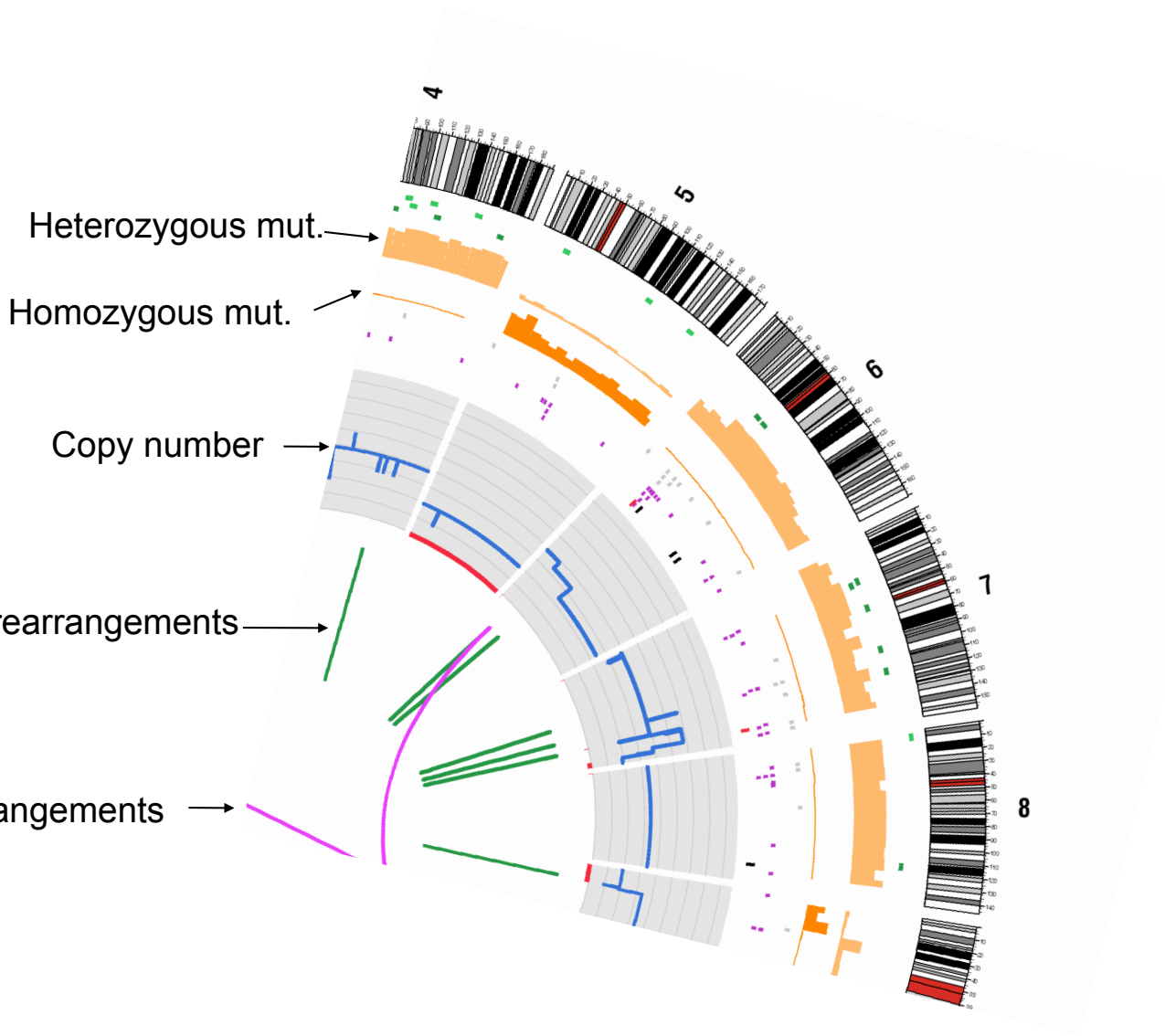




Many somatic mutations in melanoma are due to ultraviolet light exposure

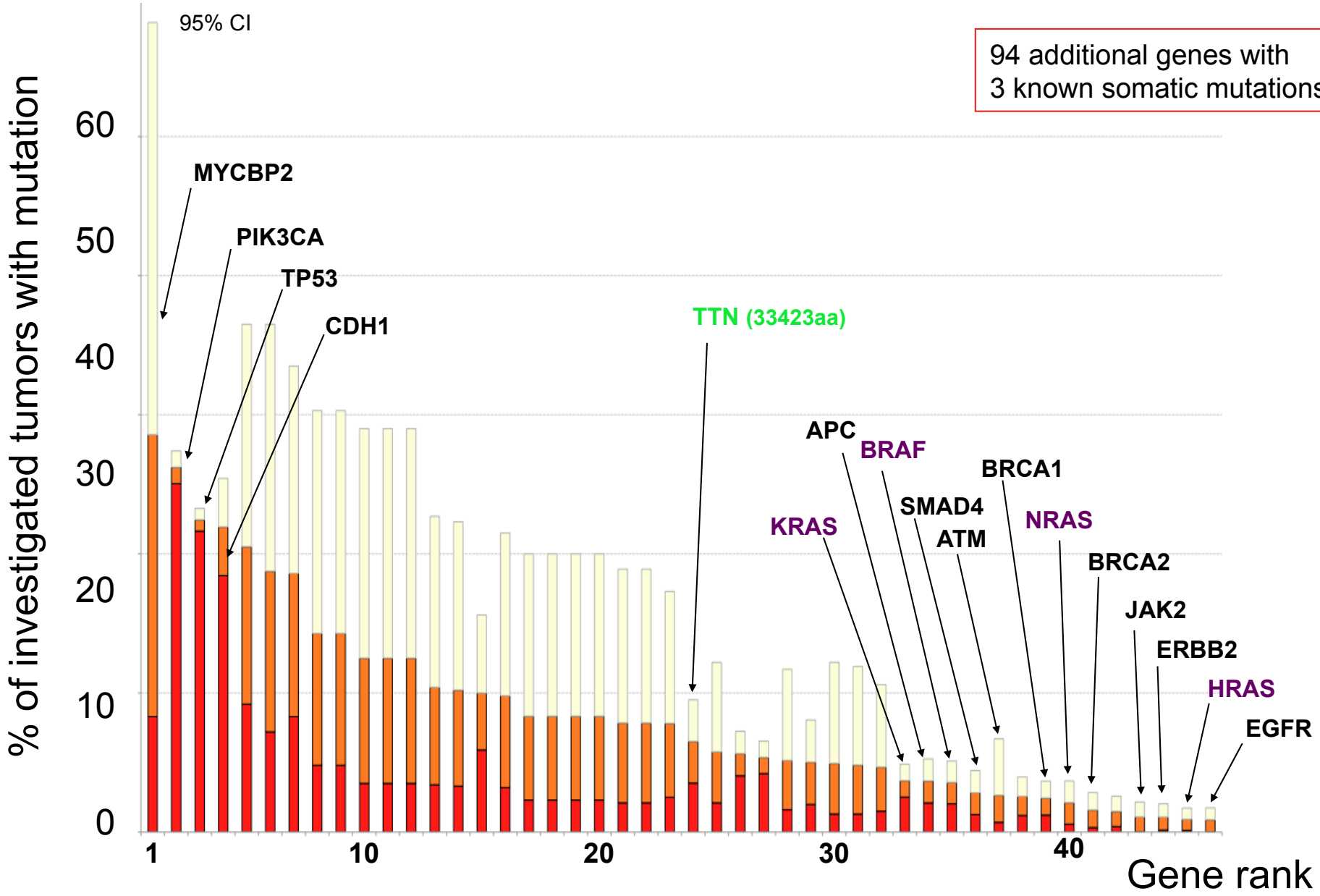





Heterozygous mut.
Homozygous mut.
Copy number
Intrachromosomal rearrangements
Interchromosomal rearrangements

"Driver" genes with at least 4 known somatic mutations in breast cancer

Cosmic Feb 2011



Conclusions from large scale screens for somatic mutation in cancer

- Frequency and type distribution of somatic mutation indicative :
  - of tissue of origin
  - of previous environmental exposure (including therapy)
  - of acquired or of innate defect in DNA repair
  - DNA repair is not uniform along the chromosomes

- For a given tumor type :
  - The set of frequently (>20%) mutated driver cancer genes is small (<10).
  - The set of rarely (<10%) mutated driver cancer genes is large (> 50) and poorly known.

- In breast cancer, on a per tumor basis
  - Several thousand somatic mutations
  - Average of 2 in-frame fusion genes
  - Average of 1 promoter fusion
  - Most of the rearranged genes are expressed

Cancer Gene Census

Genes bearing mutations causally implicated in human cancer

- 436 different genes (about 2% of all human genes)

- Type of alteration
  - Large rearrangements
    - Amplification 14 genes
    - Translocation 301 genes
    - Large deletions 33 genes
  - point mutations
    - Missense m. 118 genes
    - Nonsense m. 78 genes
    - Frame shift m. 82 genes
    - Splicing m. 52 genes

- Germline mutations 74 genes

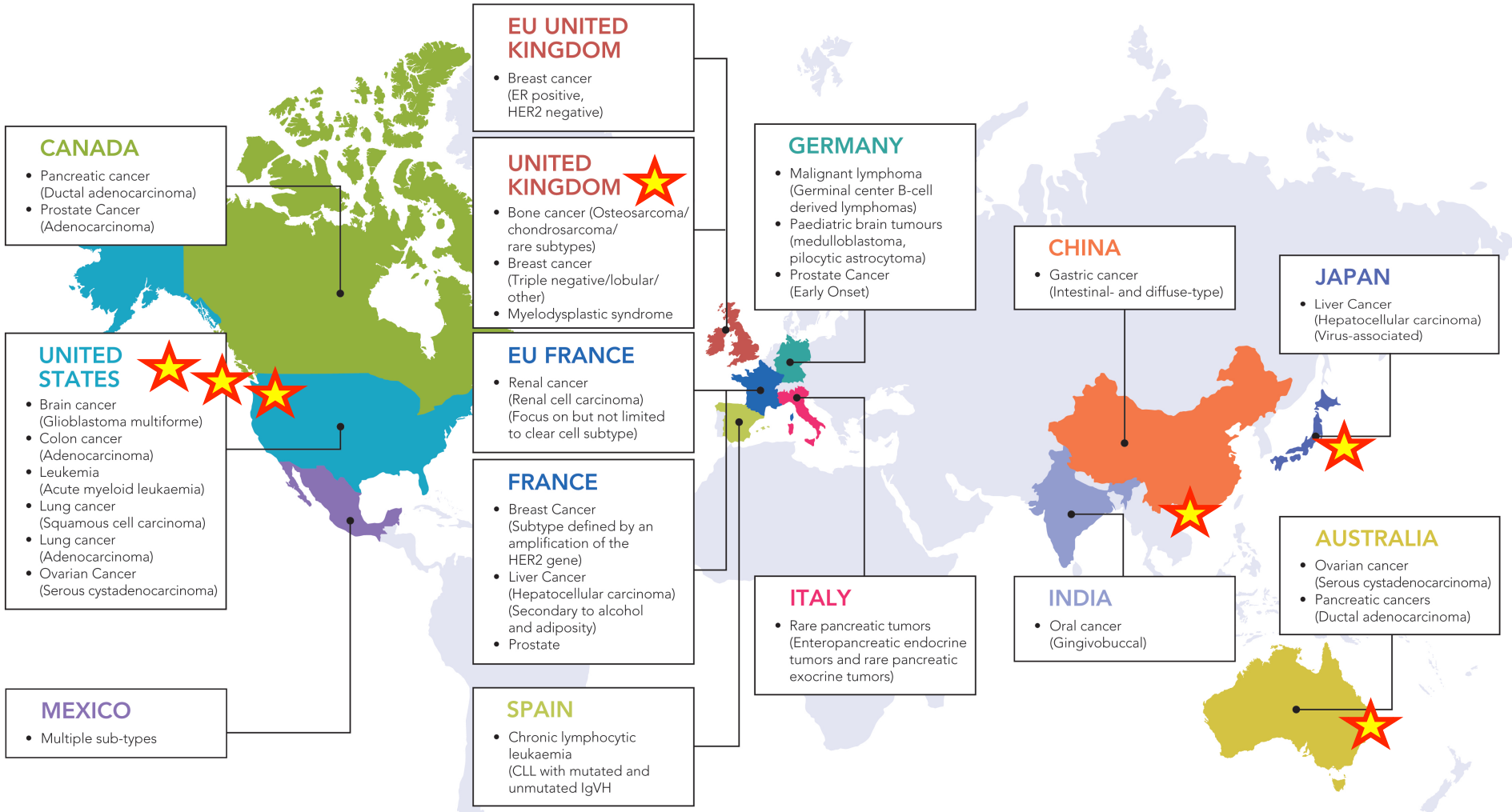
Cosmic , Feb 2011

Whole Genome shotgun sequencing

Author	Tissue	Year	Source	Pubmed
Varela *et al*	Renal Carcinoma	2011	Nature	21248752
ICGC	Liver	2010	ICGC data portal	
ICGC	Breast	2010	ICGC data portal	
ICGC	Colorectal	2010	ICGC data portal	
ICGC	Glioblastoma	2010	ICGC data portal	
ICGC	Pancreas	2010	ICGC data portal	
ICGC	Lung	2010	ICGC data portal	
TCGA	Ovarian Serous Carcinoma	2010	TCGA data portal	
Ding *et al*	Breast	2010	Nature	20393555
Campbell *et al*	Pancreas	2010	Nature	20981101
Pleasance *et al*	Lung	2010	Nature	20016488
Pleasance *et al*	Malignant Melanoma	2010	Nature	20016485
Stephens *et al*	Breast	2009	Nature	20033038
Shah *et al*	Breast	2009	Nature	19812674
Mardis *et al*	Leukemia	2009	NEJM	19657110
Campbell *et al*	Lung	2008	Nature Genetics	18438408

Cosmic , Feb 2011

ICGC Map - Nov 2010



Present applications of new generation sequencing

- Full characterization of complex nucleic acids mixtures
  - Whole genome
  - Whole transcriptome

- Targeted sequencing
  - Long range PCR
    - -Specific genes (e.g. BRCA1) or families of genes (e.g. tyrosine kinases)
  - Hybrid Capture (Pull down):
    - All exons (Exome)
    - All CpG islands

- Chromatin conformation/interaction