Genome Biology of Cancer

Rolf I. Skotheim
October 31st, 2011

Genome Biology Group
Dept. Cancer Prevention
Inst. Cancer Research
Norwegian Radium Hospital

Oslo University Hospital
Norwegian Radium Hospital

Cancer Biomedicine
University of Oslo

www.rr-research.no/cancerprevention

The birth of cancer genomics

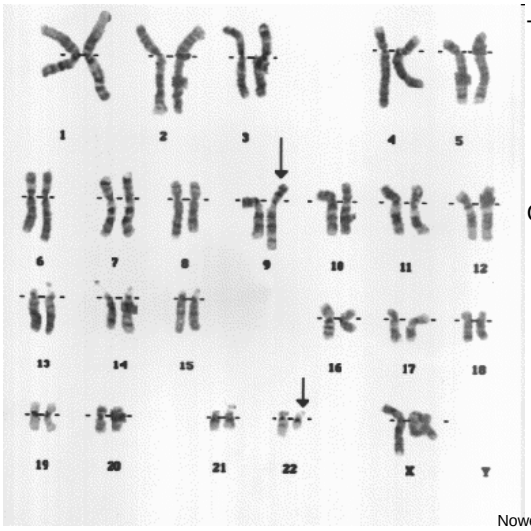
Theodor Boveri

Somatic mutation theory of cancer

Based on observations of abnormal growth of sea-urchin eggs that carry the “wrong” chromosomal complement, Boveri proposed that tumour growth is based on a similar, but particular, incorrect combination of chromosomes.

Boveri TH (1914). Zur Frage der Entstehung maligner Tumoren. Verlag von Gustav Fisher, Jena.

Karyotype with the Philadelphia chromosome



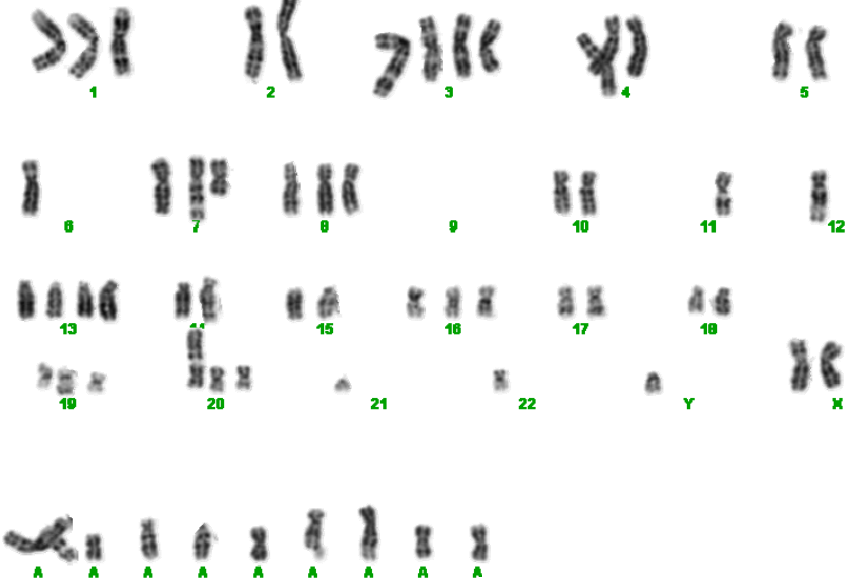
Translocation between chromosomes 9 & 22 in a chronic myelogenous leukaemia (CML); t(9;22)(q34;q11).

Giving rise to the *BCR1-ABL* fusion gene.

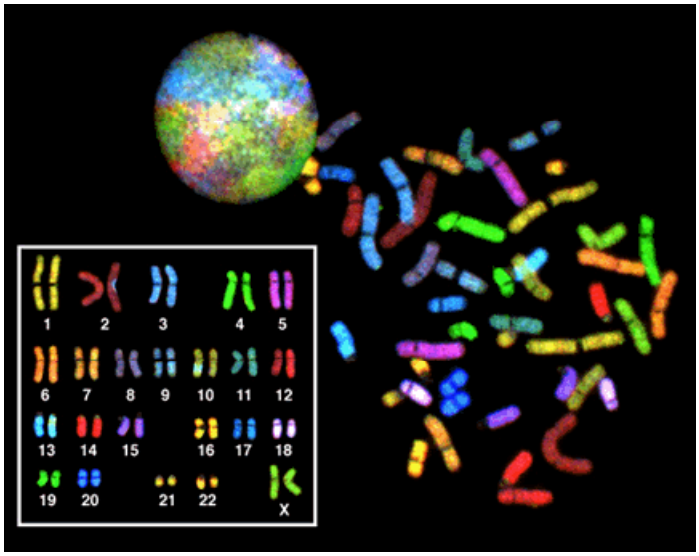
Nowell & Hungerford. Chromosomes of normal and leukemic human leucocytes. J. Natl. Cancer Inst. 1960

Complex karyotype (testicular cancer)

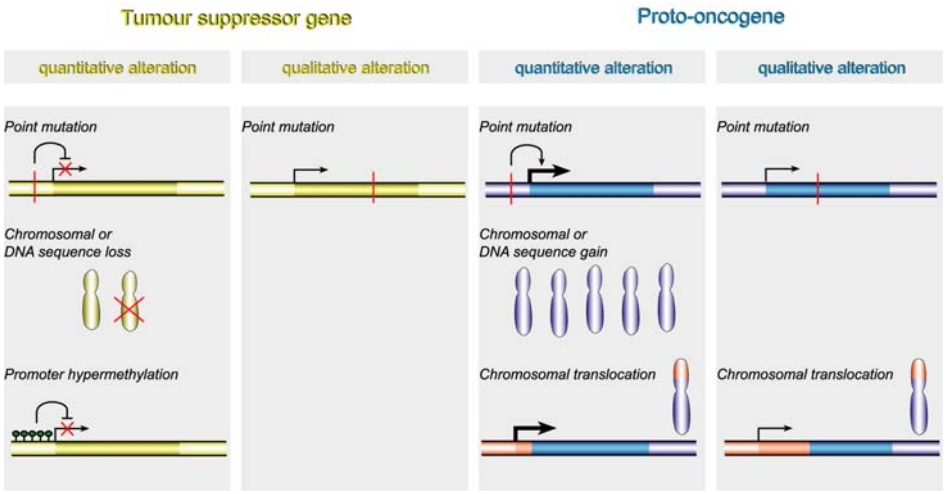
Complex karyotype (testicular cancer)



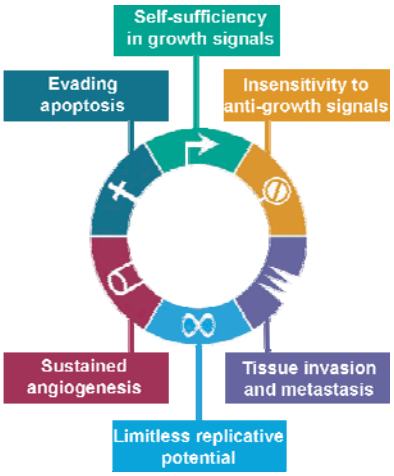
Spectral karyotype, the metaphase genome



Changes that activate proto-oncogenes and inactivate tumour suppressor genes

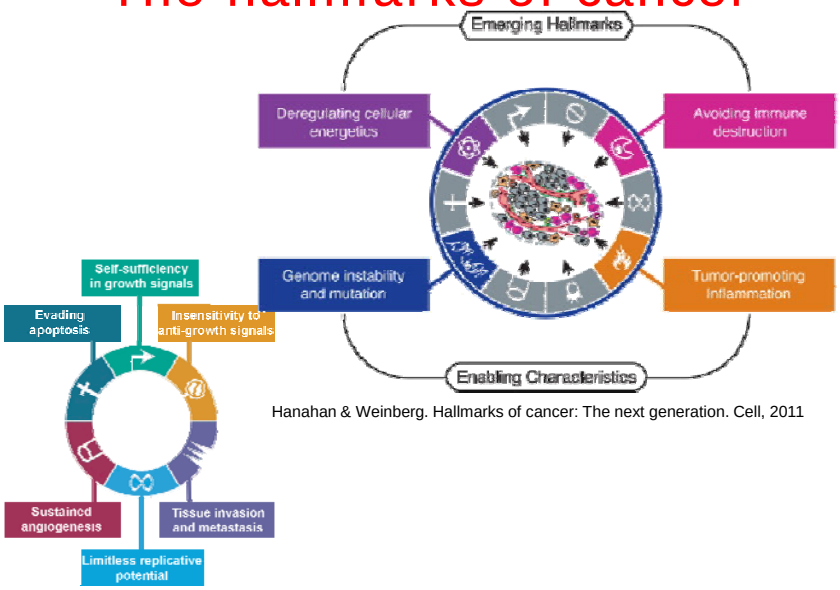


The hallmarks of cancer



Hanahan & Weinberg. The hallmarks of cancer. Cell, 2000

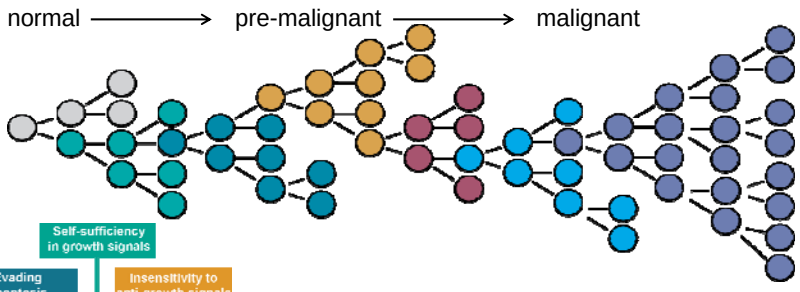
The hallmarks of cancer



Hanahan & Weinberg. Hallmarks of cancer: The next generation. Cell, 2011

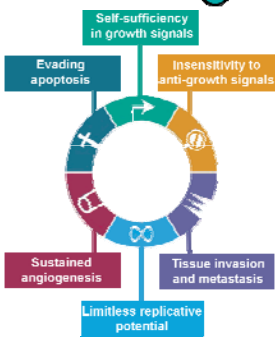
Hanahan & Weinberg. The hallmarks of cancer. Cell, 2000

Clonal selection



Nowell. The clonal evolution of tumor cell populations. Science, 1976

Important concepts:
Driver *vs.* passenger mutations



Hanahan & Weinberg. The hallmarks of cancer. Cell, 2000

Tumour suppressor genes

- Protect against malignant phenotype.
- Knudson's two-hit model for inactivation of tumour suppressor genes.

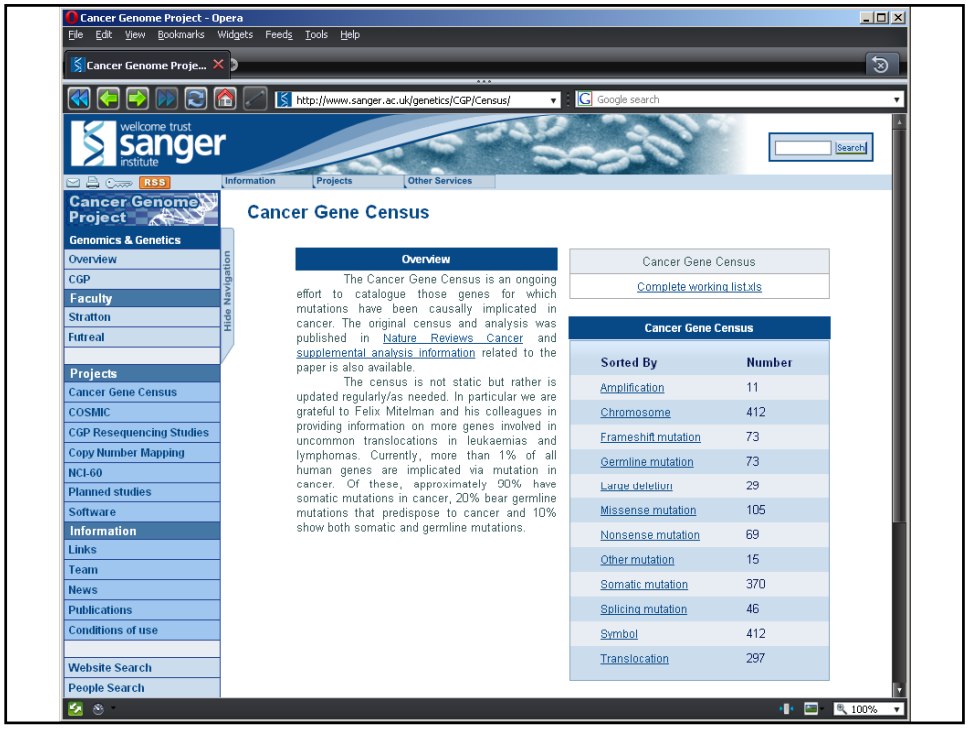
Knudson. Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA, 1971

- Examples of tumour suppressor genes:
  - *RB1, TP53, CDKN2A, PTEN, APC*
- Classes of tumour suppressors:
  - Gatekeepers vs. caretakers
  - Mutations in caretaker genes (*e.g.* DNA repair genes) lead to genomic instability, which again increases likelihood of mutations in gatekeeper genes (and in proto-oncogenes)

Proto-oncogenes

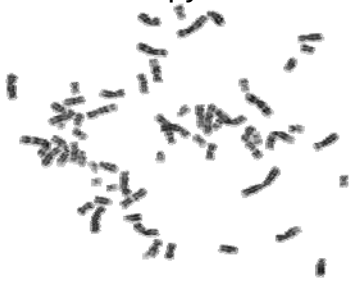
Conversion from proto-oncogene to oncogene is dominant, and oncogenes get hyperactive through quantitative or qualitative changes.

- Quantitative: overexpression
  - Gene amplification (*ERBB2*)
  - Chromosomal translocation (*IGH-MYC*)
  - Point mutations with regulatory effects (*e.g.* in promoter, UTRs)
  - Trans effects: transcription factors, cell signalling, viral integration, etc.
- Qualitative: functional switch
  - Gain-of-function mutation (*e.g.* point mutation; *KRAS, EGFR*)
  - Chromosomal translocation (*BCR-ABL1*)
  - Alternative splicing (*BCL2L1*)



Strategies for identification of proto-oncogenes

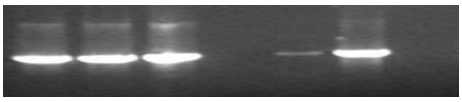
- Transfections of parts of cancer genomes into immortalized mouse-derived fibroblastic cells (gain-of-function; *e.g. RAS)*
- Cytogenetics (*e.g.* translocations and amplifications)
- RNAi, causing *e.g.* reduced cell growth or increased apoptosis
- Microarrays/HT-sequencing (combining measurements of DNA copy numbers with gene expression)



Use of genome-level tools in identification of cancer-critical genes

- Gene expression microarrays
  - gene level
  - exon level
  - fusion transcripts
- DNA copy number microarrays
- Tissue microarrays
- Functional cell microarrays
- High-throghput sequencing (DNA & RNA)

DNA microarrays, Results and visualisation



samples

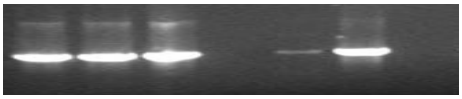
GeneNo	sample 1	sample 2	sample 3	sample 4	sample 5	sample 6	sample 7
Gene_1	0.919	0.836	0.795	1.026	0.553	0.788	1.003
Gene_2	2.861	2.483	2.101	1.369	4.683	5.299	0.920
Gene_3	1.582	1.851	1.657	1.144	1.835	1.500	1.043
Gene_4	1.197	1.267	1.287	0.843	0.583	0.870	1.413
Gene_5	0.780	0.695	0.844	1.082	0.540	0.820	3.099
Gene_6	0.745	0.974	0.829	0.735	2.678	0.906	1.073
Gene_7	1.191	0.794	1.175	1.111	0.606	0.688	1.233
Gene_8	1.748	2.741	1.958	0.695	2.079	1.501	0.971
Gene_9	1.118	1.613	0.808	1.311	1.050	1.000	0.756
Gene_10	0.974	1.000	0.900	0.839	0.627	1.196	1.523
Gene_11	0.970	0.865	0.993	1.075	1.202	0.995	1.052
Gene_12	0.731	0.722	0.622	0.817	0.736	0.774	1.000
Gene_13	0.988	1.197	1.144	1.183	0.776	0.718	1.389
Gene_14	1.012	1.072	0.802	1.004	1.325	1.395	1.212
Gene_15	1.625	1.340	2.034	2.008	0.618	0.961	0.873
Gene_16	1.336	2.592	2.164	1.158	0.409	2.510	1.223
Gene_17	4.201	5.179	5.924	0.689	1.624	2.729	0.839
Gene_18	0.946	0.571	0.923	1.522	1.571	1.458	0.371
Gene_19	1.432	1.286	1.555	1.385	0.696	0.890	0.982
Gene_20	1.002	0.790	1.107	1.252	0.594	0.577	0.883

genes/ probes

sample intensity value / reference intensity value

➔ red/green ratio

DNA microarrays, Results and visualisation



samples

GeneNo	sample 1	sample 2	sample 3	sample 4	sample 5	sample 6	sample 7
Gene_1	0.919	0.836	0.795	1.026	0.553	0.788	1.003
Gene_2	2.861	2.483	2.101	1.369	4.683	5.299	0.920
Gene_3	1.582	1.851	1.657	1.144	1.835	1.500	1.043
Gene_4	1.197	1.267	1.287	0.843	0.583	0.870	1.413
Gene_5	0.780	0.695	0.844	1.082	0.540	0.820	3.099
Gene_6	0.745	0.974	0.829	0.735	2.678	0.906	1.073
Gene_7	1.191	0.794	1.175	1.111	0.606	0.688	1.233
Gene_8	1.748	2.741	1.958	0.695	2.079	1.501	0.971
Gene_9	1.118	1.613	0.808	1.311	1.050	1.000	0.756
Gene_10	0.974	1.000	0.900	0.839	0.627	1.196	1.523
Gene_11	0.970	0.865	0.993	1.075	1.202	0.995	1.052
Gene_12	0.731	0.722	0.622	0.817	0.736	0.774	1.000
Gene_13	0.988	1.197	1.144	1.183	0.776	0.718	1.389
Gene_14	1.012	1.072	0.802	1.004	1.325	1.395	1.212
Gene_15	1.625	1.340	2.034	2.008	0.618	0.961	0.873
Gene_16	1.336	2.592	2.164	1.158	0.409	2.510	1.223
Gene_17	4.201	5.179	5.924	0.689	1.624	2.729	0.839
Gene_18	0.946	0.571	0.923	1.522	1.571	1.458	0.371
Gene_19	1.432	1.286	1.555	1.385	0.696	0.890	0.982
Gene_20	1.002	0.790	1.107	1.252	0.594	0.577	0.883

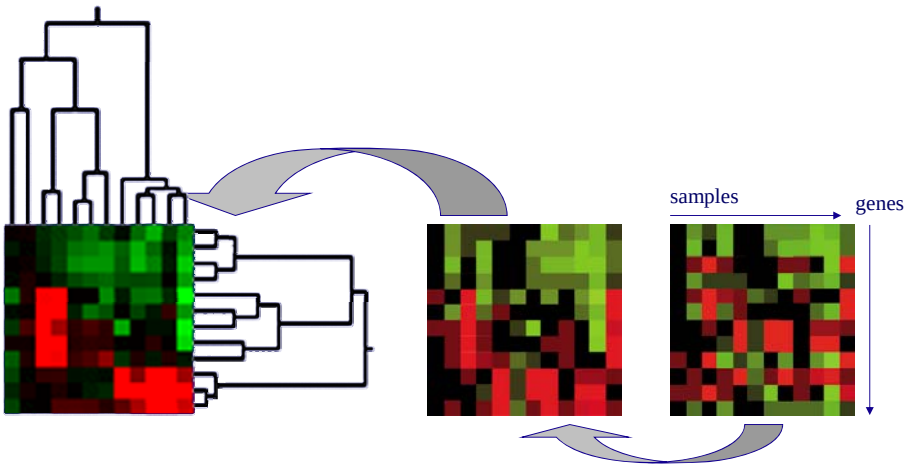
genes/
probes

sample intensity value
reference intensity value

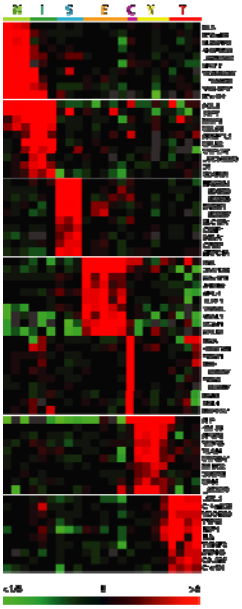
➔ red/green ratio

DNA microarrays, Results and visualisation

- **Hierarchical cluster analysis**

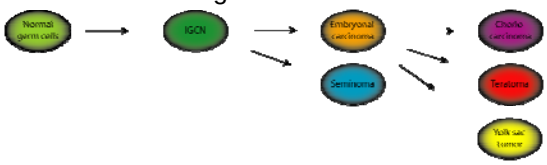


DNA microarrays, Results and visualisation

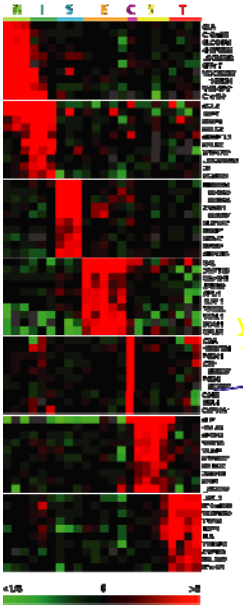


- **Hierarchical cluster analysis**
- **Significance Analysis of Microarrays (SAM)**
  - individual analyses of each subtype against the rest of the sample set

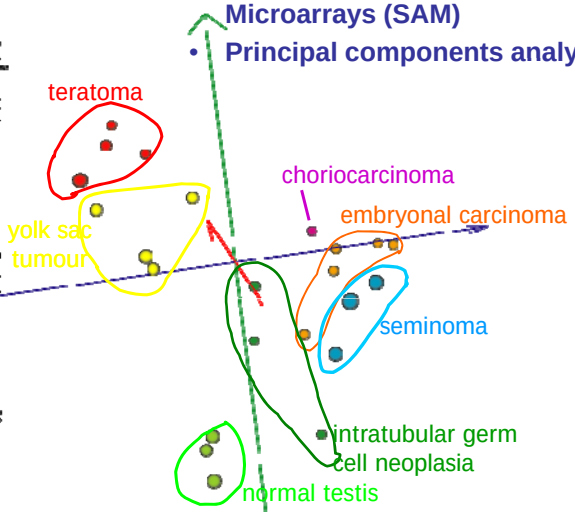
Testicular tumourigenesis



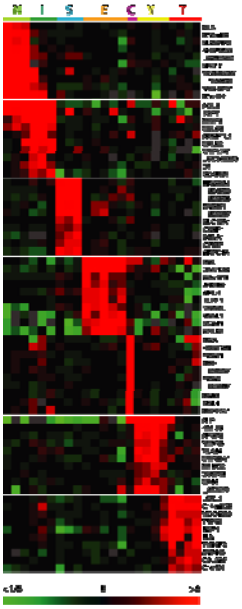
DNA microarrays, Results and visualisation



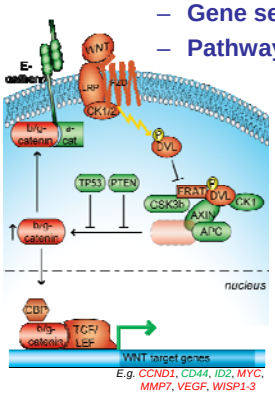
- **Hierarchical cluster analysis**
- **Significance Analysis of Microarrays (SAM)**
- **Principal components analysis**



DNA microarrays, Results and visualisation



- **Hierarchical cluster analysis**
- **Significance Analysis of Microarrays (SAM)**
- **Principal components analysis**
- **Systems biology tools**
  - **Gene sets enrichments analysis**
  - **Pathway mapping**

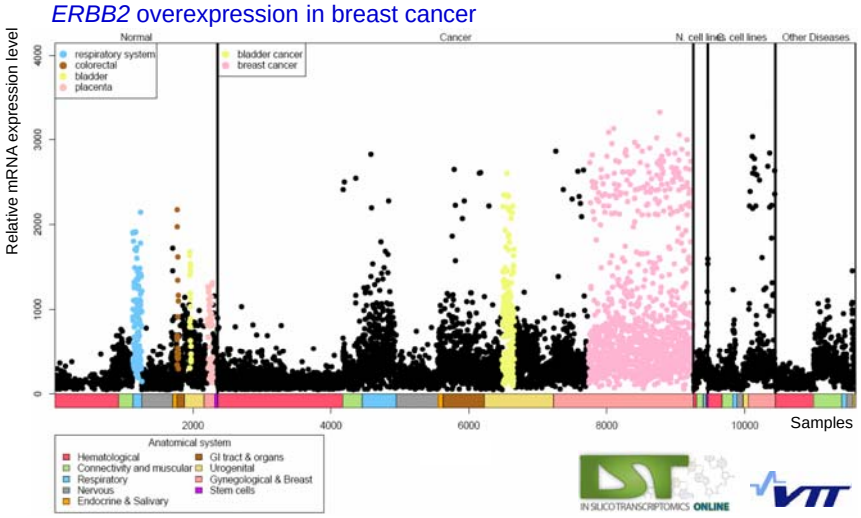


Gene expression databases

- Data repositories:
  - Gene expression omnibus (GEO) at the National Center for Biotechnology Information (NCBI): www.ncbi.nlm.nih.gov/geo
  - ArrayExpress at the European Bioinformatics Institute (EBI)
- Database interpretation and visualisation
  - Oncomine (www.oncomine.org)
  - MediSapiens (www.medisapiens.com)

MediSapiens, visualisation of gene expression across the human sample-space

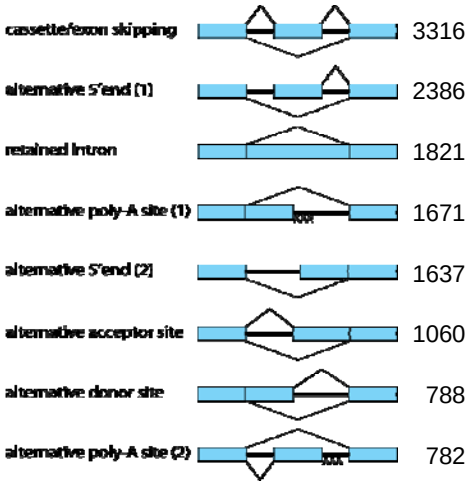
ERBB2 overexpression in breast cancer



Herceptin, an antibody against ERBB2 (alias HER2), used to treat breast cancers with amplification of the *ERBB2* gene (17q12)

Kilpinen *et al.*, Genome Biology, 2008

Transcript variation



3316
2386
1821
1671
1637
1060
788
782

Other: 2429

The most common sources of variation among human transcripts (mostly alternative splicing)

Most expression microarray platforms provide one measurement per gene

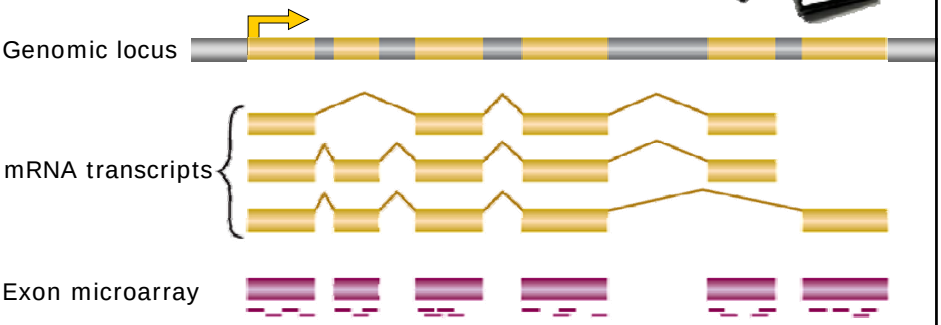
Statistics from ASTRA (Alternative Splicing and TRanscription Archives)
http://alterna.cbrc.jp/index.php

Affymetrix exon microarrays

Provides gene expression measurements from individual exons

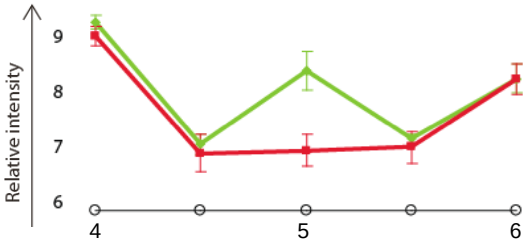
5.4 million probes, measuring

- 325 000 known exons
- ~1 million extra exons based on sequence prediction

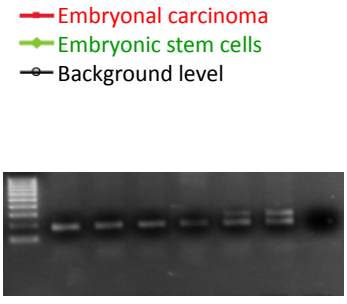


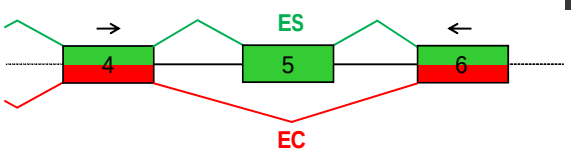
Most DNA microarrays provide one single measurement per gene (in the 3' end)

Detection of alternative splicing events

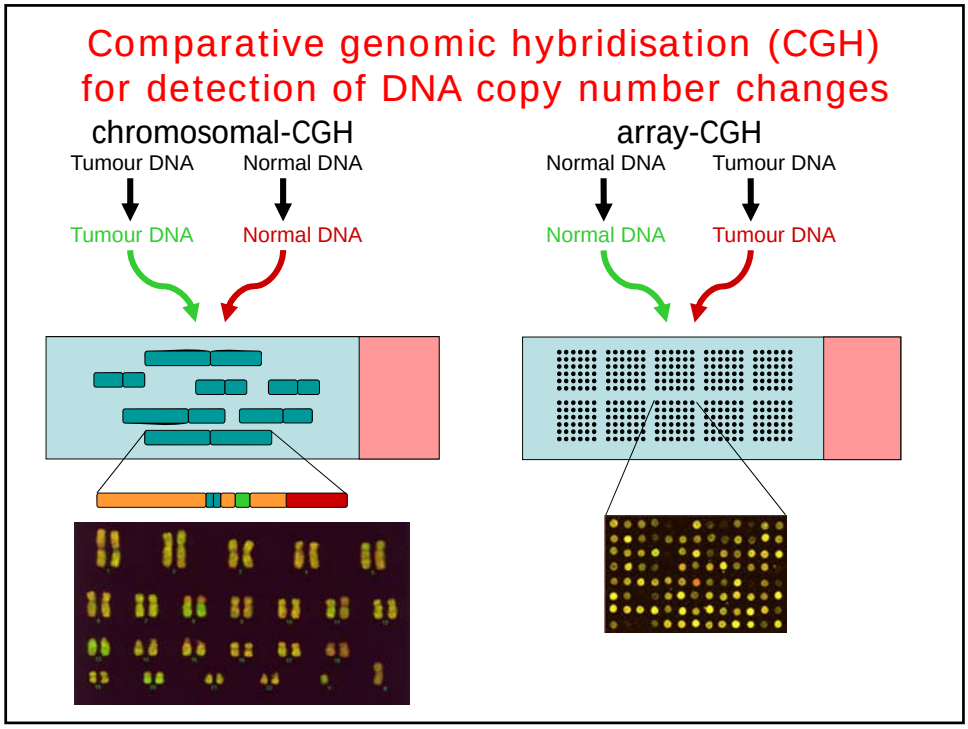


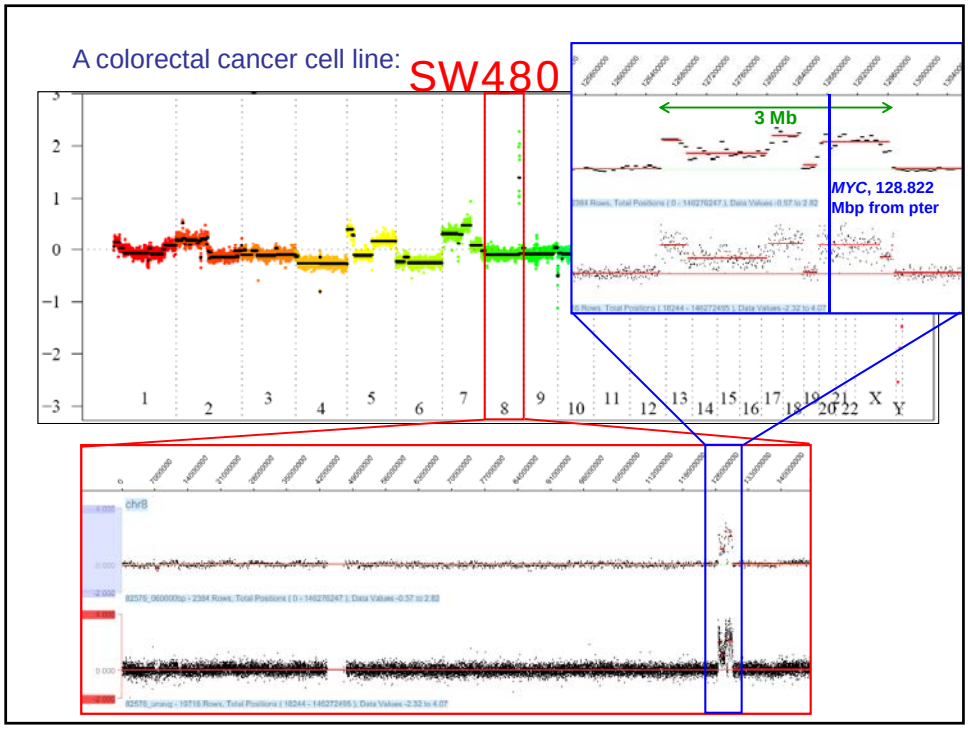
Probes in 5' to 3' direction within a gene encoding a **zink finger transcription factor**



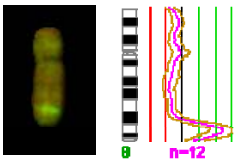


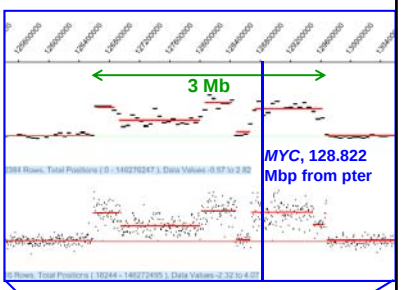
Alagaratnam *et al.*, in prep.

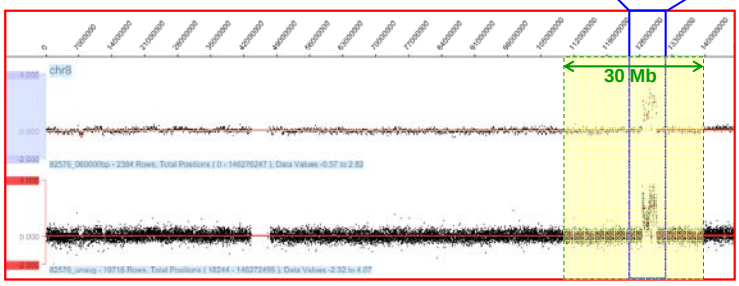

Comparative genomic hybridisation (CGH)
for detection of DNA copy number changes
chromosomal-CGH
Tumour DNA
Normal DNA
Tumour DNA
Normal DNA
array-CGH
Normal DNA
Tumour DNA
Normal DNA
Tumour DNA


A colorectal cancer cell line:
SW480
3 Mb
MYC, 128.822
Mbp from pter
chr8

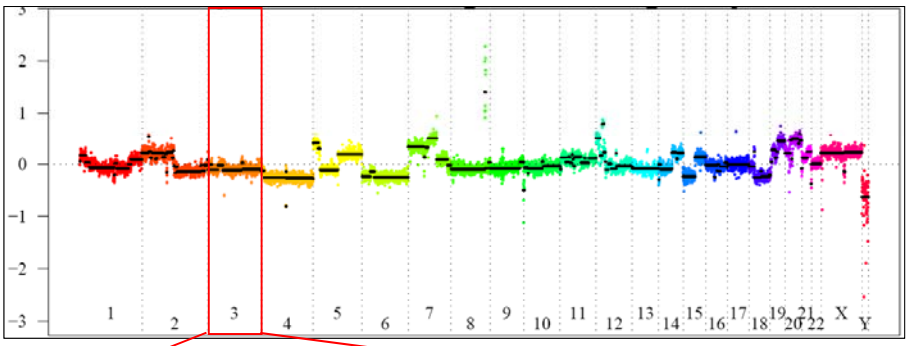
A colorectal cancer cell line: SW480



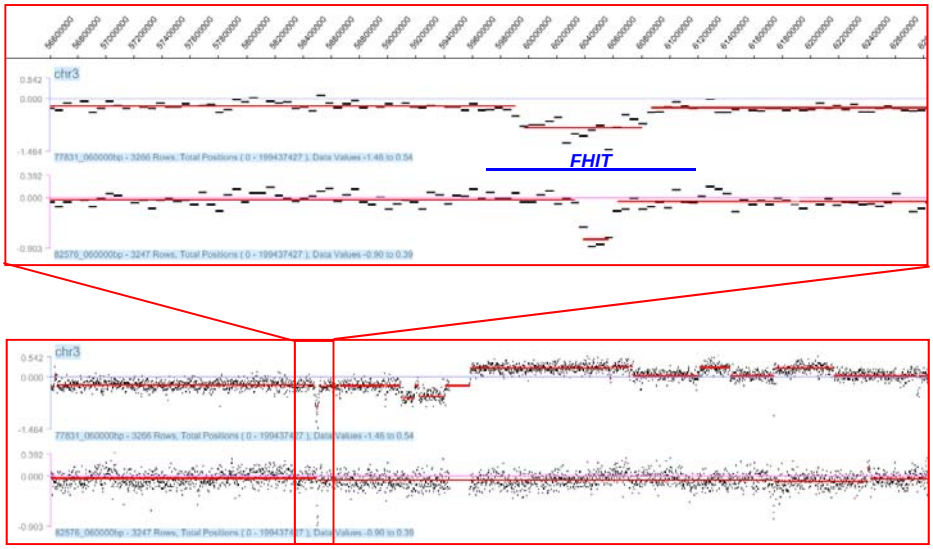




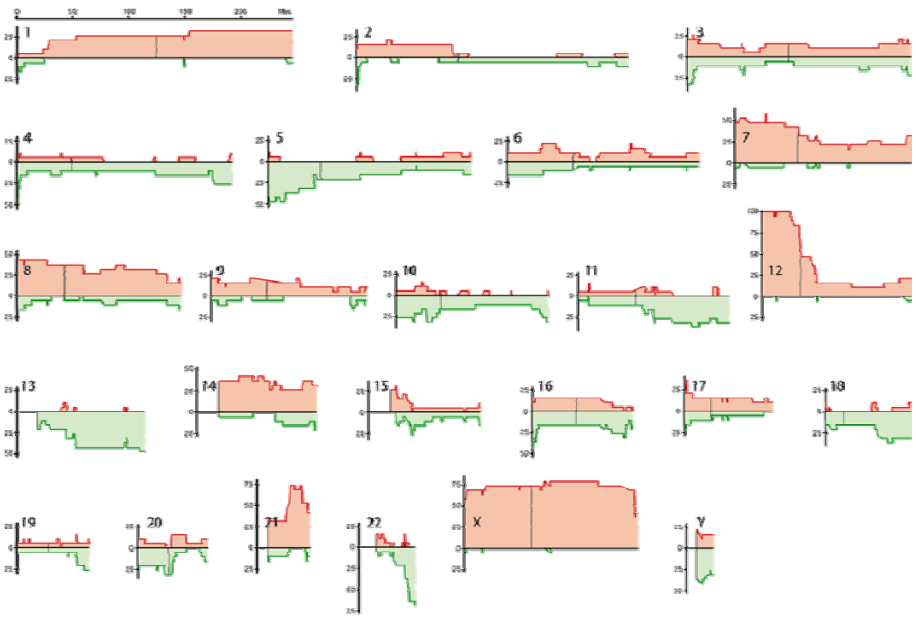
A colorectal cancer cell line: SW480



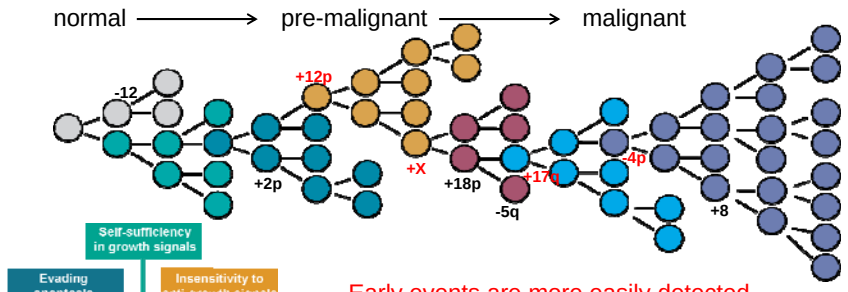
HT29 & SW480, chr. 3

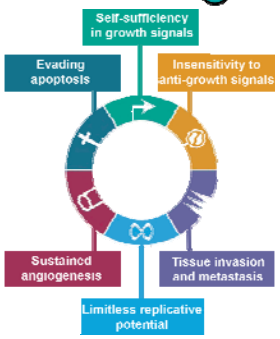


Testicular cancer genome profile (array-CGH)



Clonal selection





Hanahan & Weinberg. The hallmarks of cancer. Cell, 2000

Early events are more easily detected

Important events are more likely to be non-random across a series of cancers. This is because these are events which give rise to selective advantage

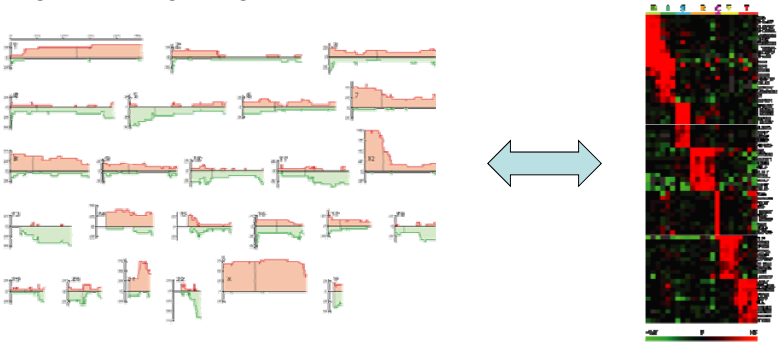
The selective advantage is probably due to one or several genes which expression are altered as a consequence of the DNA copy number change

Integrated genome and transcriptome analyses

Identification of genes that are over- or under-expressed in samples also having increased or decreased DNA copy number of these genes' loci.

➔ potential **proto-oncogenes** or **tumour suppressor genes**

i.e., genes that give the cells selective advantage from having their genomic regions gained or lost.

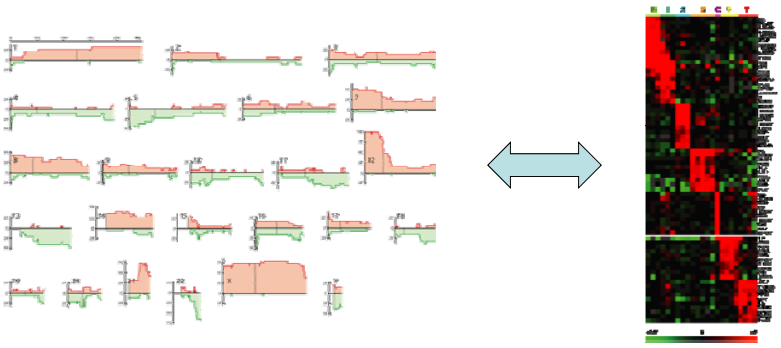


Integrated genome and transcriptome analyses

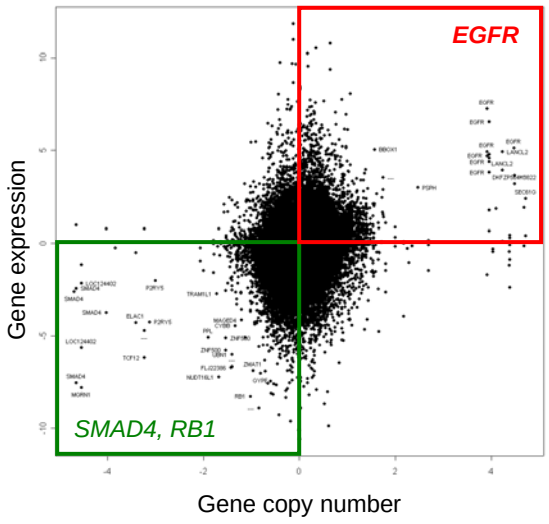
For each sample, each gene have two measurements:

- DNA copy number level
- RNA expression data

➔ Crude analysis: plot these against each other!



DNA-RNA outlier analysis

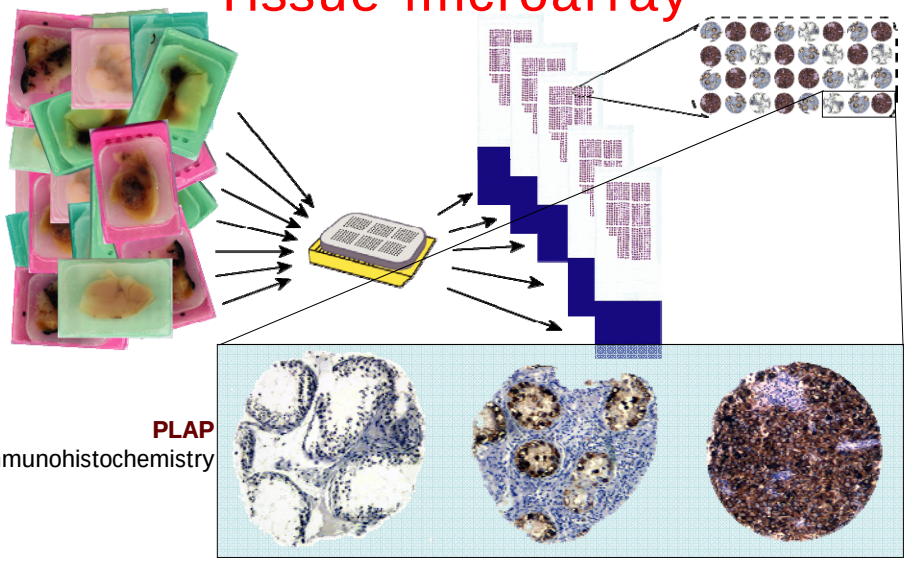


Integration of gene copy numbers with gene expression levels in the same sample

Clear segregation of known key cancer genes, along with a handful of novel genes

Figure from Henrik Edgren *et al.*

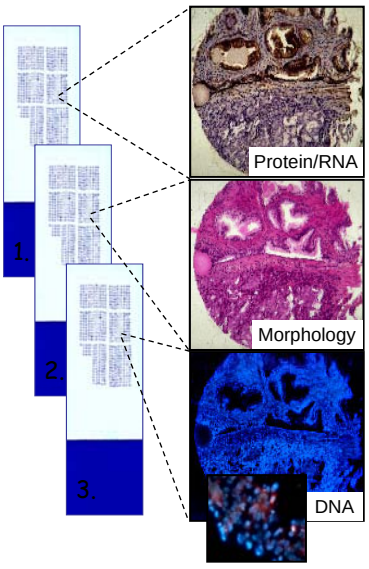
Tissue microarray



PLAP
immunohistochemistry

Normal testis | premalignant | Seminom

Tissue microarray - Applications



Immunohistochemistry
RNA *in situ* hybridisation

Hematoxylin and eosin

Fluorescence *in situ* hybridisation

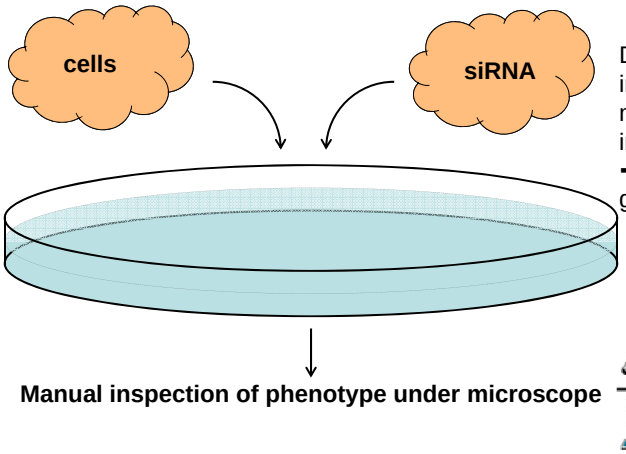
Tissue microarray - Benefits

- Analyse many tissue samples simultaneously
- Standardisation
  - Minimises experimental errors caused by differing conditions
- "Amplification" of tissue resources
- Saves you time, money, and labour

Applications in cancer research

- Does your candidate gene/protein have prognostic value?
- Can it be used as a diagnostic marker?
- What is the frequency of a molecular alteration in different tumour types?

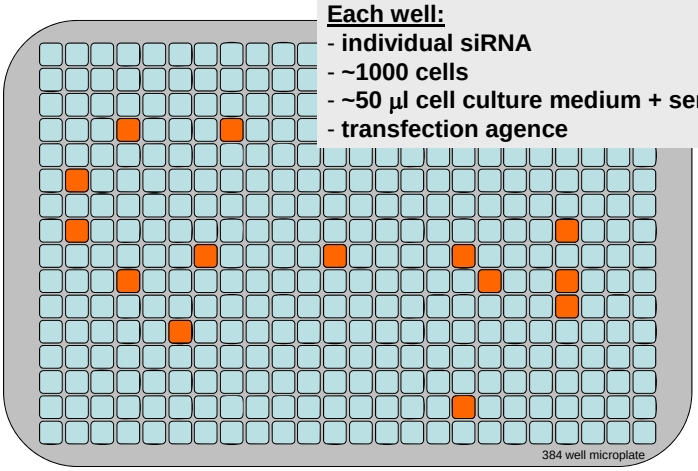
siRNA, knockdown of gene expression



Double-stranded short interferring RNA (siRNA) molecule to induce RNA-interference (RNAi)
➔ specific knockdown of gene expression

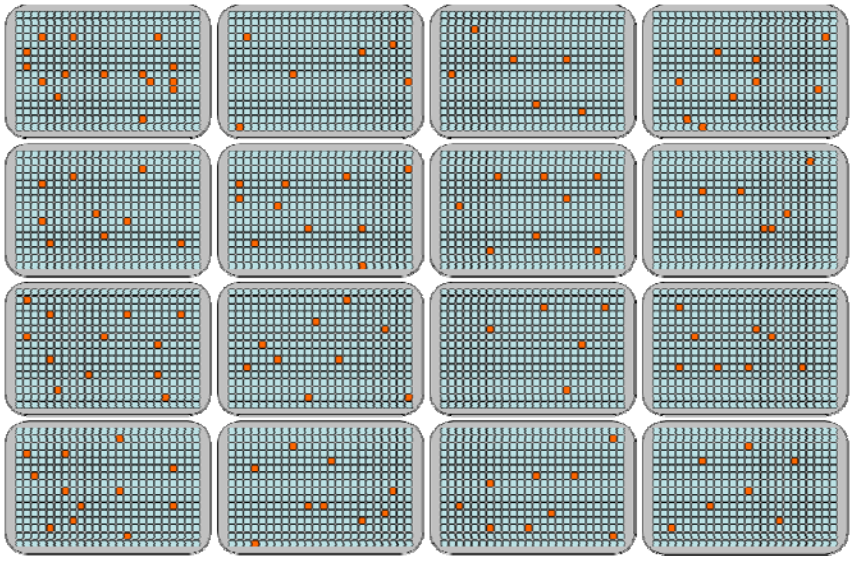
High-throughput siRNA screening

Each well:
- **individual siRNA**
- **~1000 cells**
- **~50 µl cell culture medium + serum**
- **transfection agence**

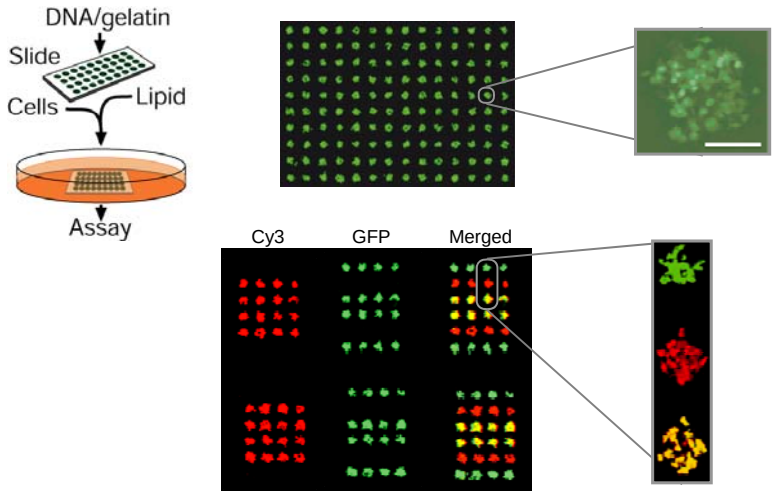


High-throughput siRNA screening

RNAi goes genomic: elucidating gene function with siRNA libraries



Functional cell microarray

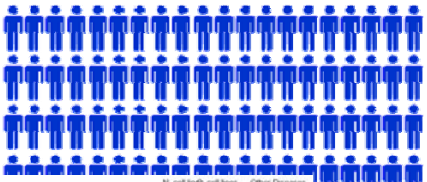


Ziauddin & Sabatini, Nature, 2001

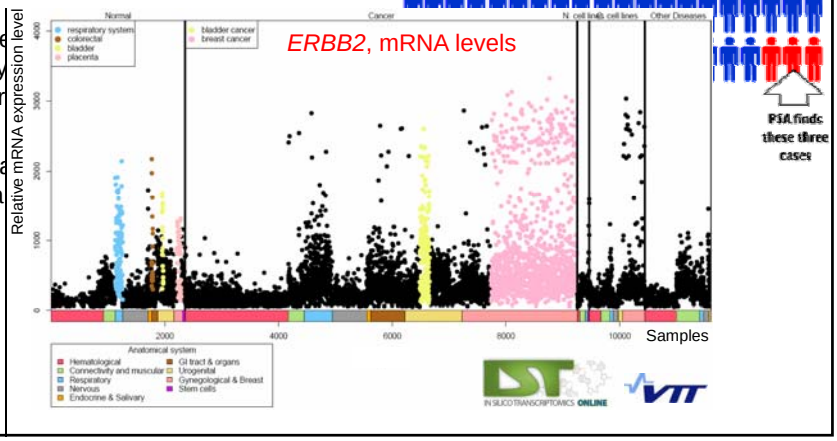
Cancer Biomarkers

Lessons from quantitative gene/protein expression

- Genes/proteins highly expressed in cancer are commonly used as biomarkers for diagnostics (*e.g.* PSA in prostate cancer) and drug sensitivity (*e.g.* ERBB2, aka HER2, in breast ca.).

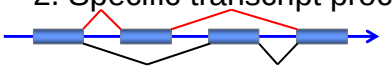


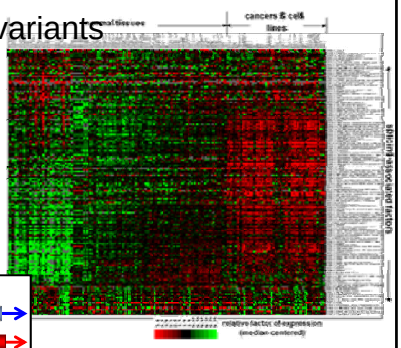
- Challe
  Alway
  requir
- Bioma
  warra



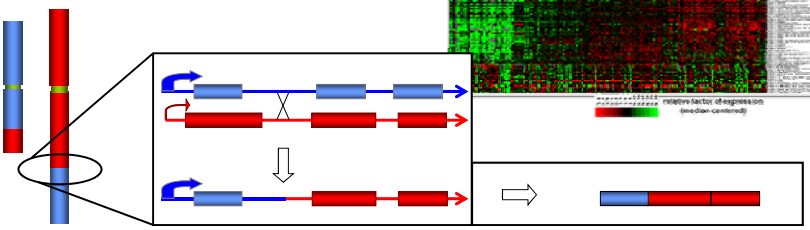
Strategies to identify novel cancer-specific transcripts

- 1. Embryo-specific transcripts
- 2. Specific transcript processing variants



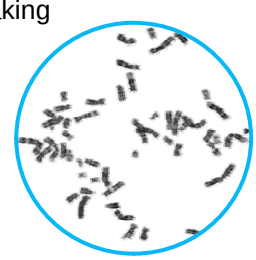


- 3. Fusion genes



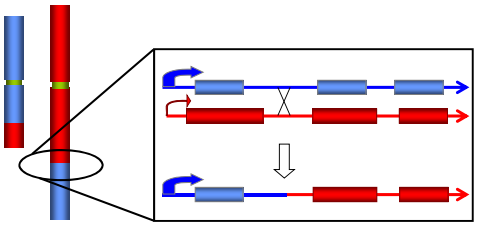
Fusion genes in cancer

- Caused by *e.g.* chromosomal translocations, deletions, and inversions
- Particularly common in haematological cancers, sarcomas, and also in certain carcinomas (*e.g.* prostate, lung, thyroid)
- Identification of certain fusion genes are currently performed for differential diagnosis or therapeutic decision-making
- Several technological limitations

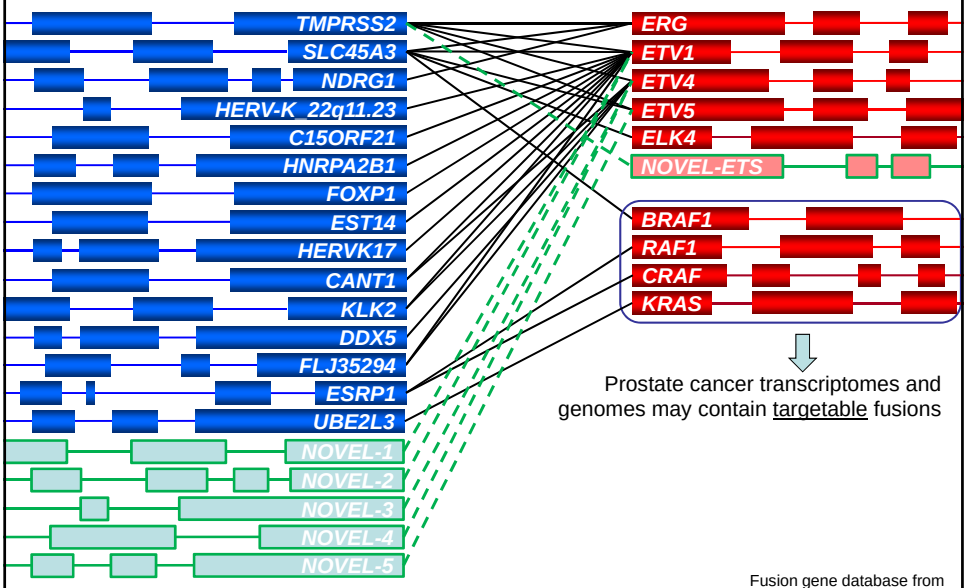


Translocations and fusion genes

- Chronic myelogenous leukaemia, t(9;22)(q34;q11)
  - Philadelphia chromosome, first identified human translocation
  - *BCR-ABL1*, encoding a fusion protein with domains from both original genes, including a tyrosine kinase (TK) activity.
  - Gleevec binds at the kinase domain, and inhibits phosphorylation of TK target proteins.
- Burkitt's lymphoma, t(8;14)(q24;q32)
  - *MYC* proto-oncogene juxtaposed with the immunoglobulin heavy chain gene: *IGH-MYC*



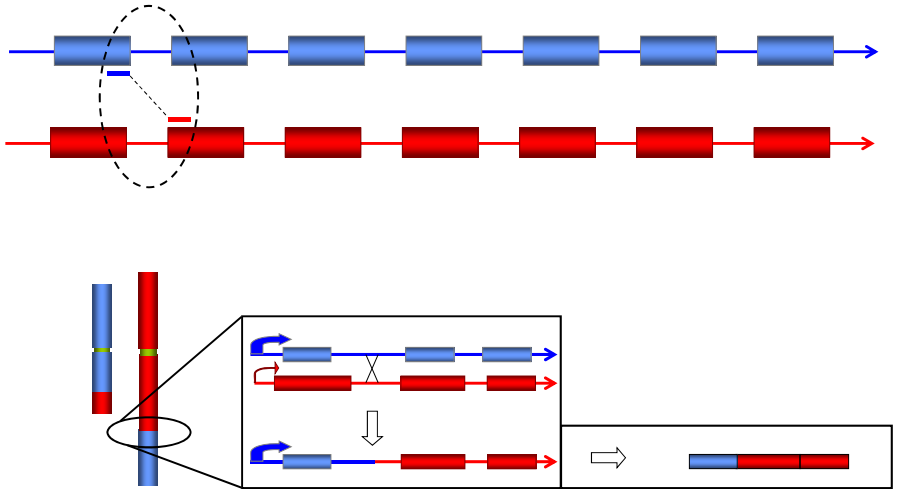
Fusion genes in prostate cancer



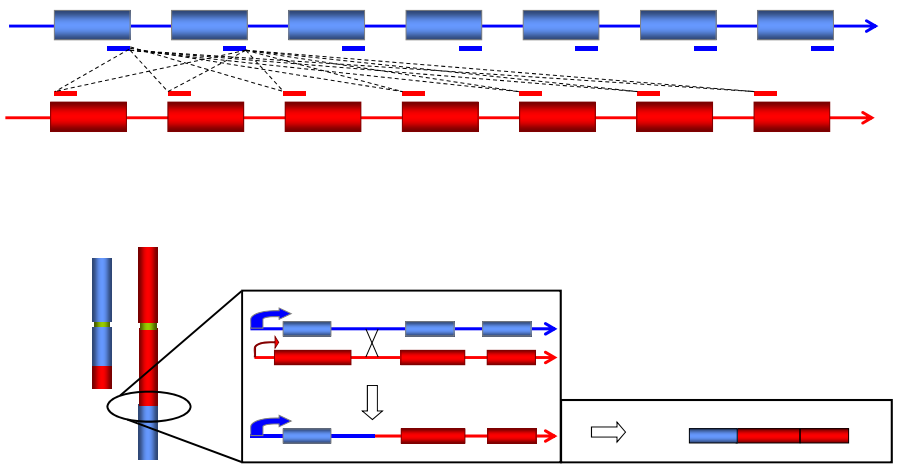
Prostate cancer transcriptomes and genomes may contain targetable fusions

Fusion gene database from
Marthe Løvf, Gard O. S. Thomassen *et al.*, Genes Chr. Cancer, 2011

Fusion gene microarray

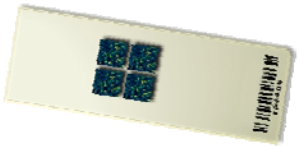


Fusion gene microarray

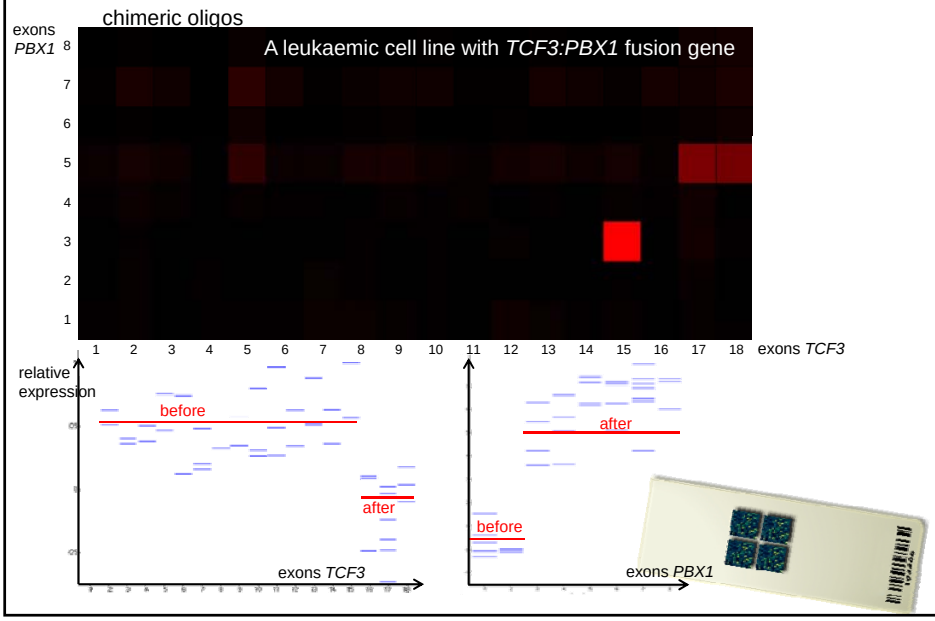


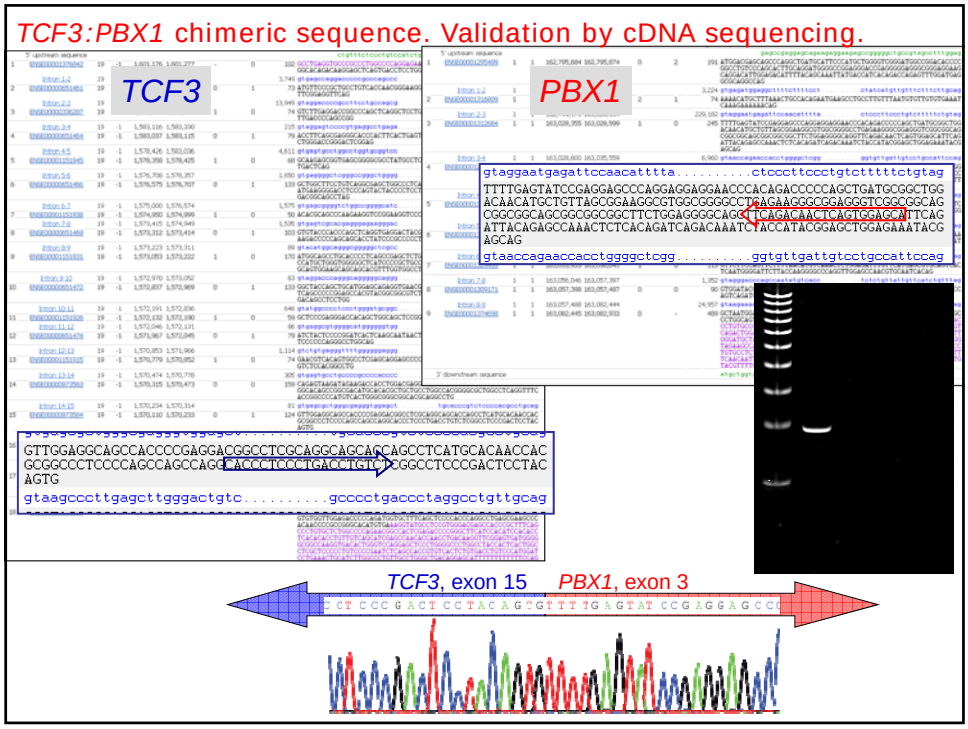
Pilot design, fusion gene microarray

- Databases/literature
  - 275 known fusion genes at time of pilot array design
- Sequences and exon annotation from Biomart.org
- Generation of chimeric sequences (~60 000)
  - Automised by script programmed in Python
- Oligo design
  - 34-40mers, variable length for isothermic properties
  - Chimeric oligos with matching Tm from up- and downstream fusion partners
  - Intragenic oligos
- Microarray platform: NimbleGen



Visualisation of results

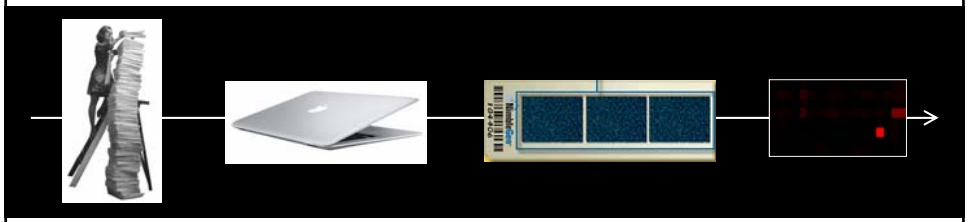




Pilot: http://www.molecular-cancer.com/content/8/1/5

Fusion gene microarray, version 2.1

- 556 gene pairs, including candidate fusion genes from high-throughput sequencing
- NimbleGen 3 x 720k microarray format
- Discovery/validation
  - Known and putative fusion genes in solid tumours
- Diagnostic potential
  - leukaemia, sarcomas, other solid tumours?

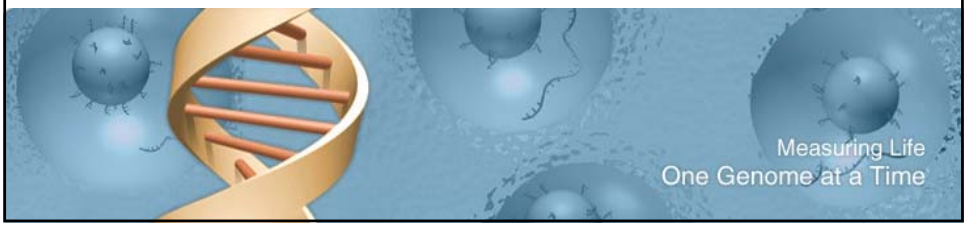


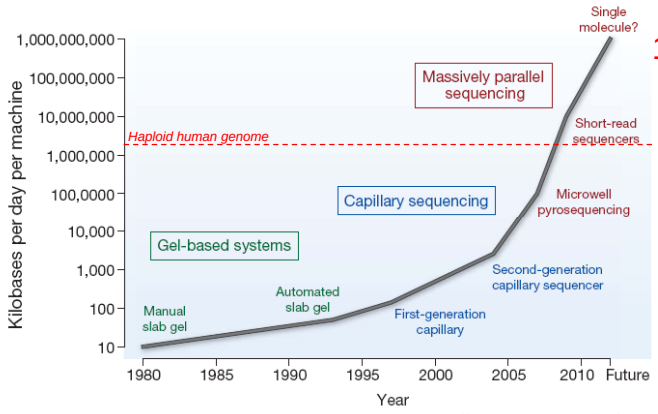
Løvf, Thomassen *et al.*, Genes Chromosomes and Cancer, 2011

Next generation sequencing

New sequencing methodologies that do not use the Sanger sequencing methodology (di-deoxy sequencing), and which have unprecedented high throughputs (giga-base level)

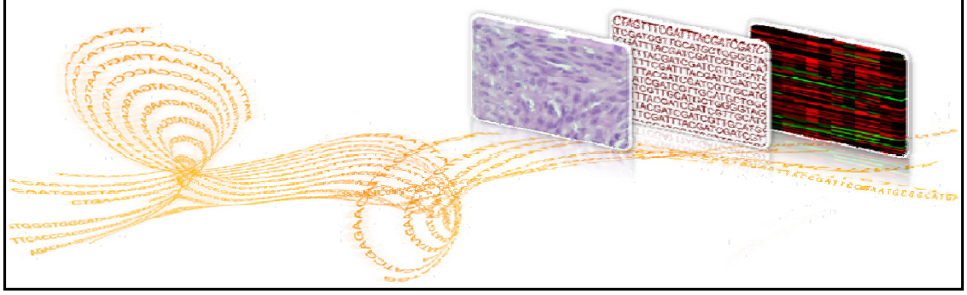
454, SOLiD, GA-IIx, HiSeq-2000, Helicos, PacBio, Ion Torrent, Dover (Polonator), & Complete Genomics





1st, 2nd, and next generations of DNA sequencing

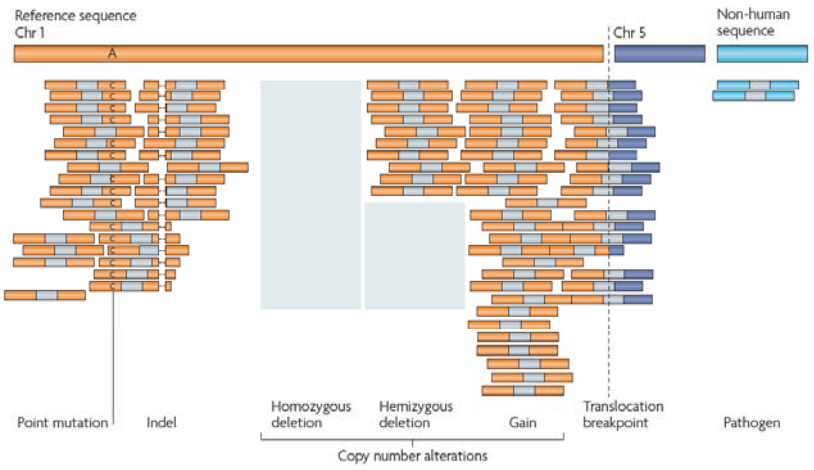
Stratton *et al.*, Nature, 2009



Main applications

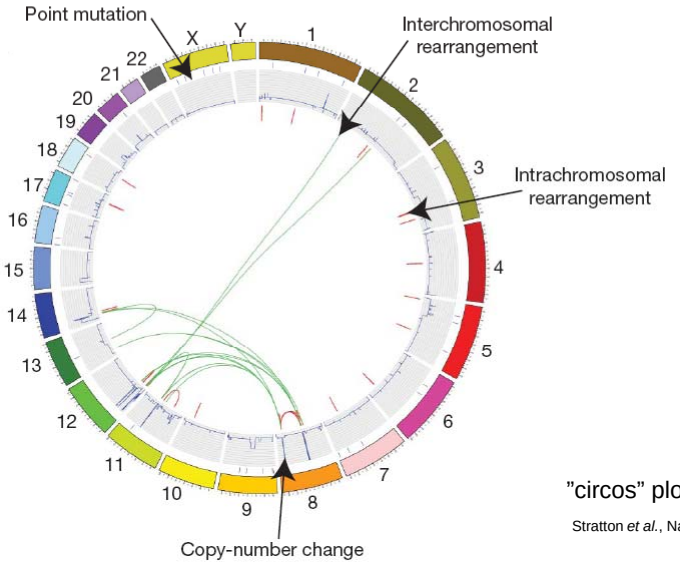
- Genome analysis, DNA-seq
  - Variant detection (mutation/polymorphism)
    - base level
    - structural changes (paired end)
  - Copy number analysis
- Transcriptome analysis, RNA-seq
  - Expression levels (tag seq)
  - Transcript variants (paired end)
  - Small RNA analysis (*e.g.* miRNA profiling)
- DNA-protein interactions
  - ChIP-seq (chromatin immunoprecipitation)
  - Transcription factor binding sites

Types of genome alterations detectable by next-generation sequencing



Meyerson *et al.*, Nat.Rev.Genet., 2010

Visualising a cancer genome



”circos” plot

Stratton *et al.*, Nature, 2009

Sequencing based transcriptome analysis (RNA-seq)

Characterize all transcriptional activity, coding and non-coding, in cell population without *á priori* assumptions

- Novel transcript structures from aberrant splicing, promoter usage, and fusion genes
- digital expression levels/relative abundance of transcripts
- mutation detection
- non-coding/regulatory RNAs

Illumina, Genome Analyzer IIx

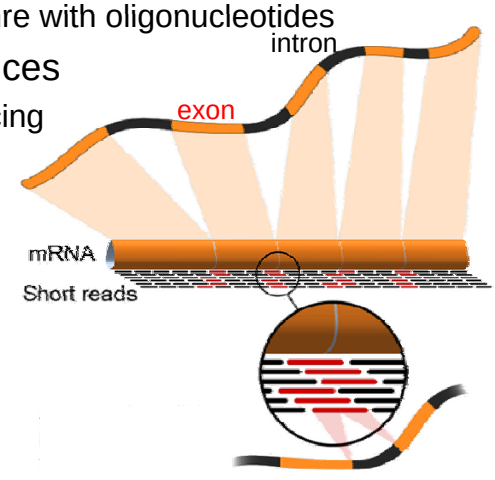
Unbiased

- ChIP-seq, not limited to known promoter areas
- RNA-seq, not limited to known transcript structures and exons
- DNA-seq, mutation analysis not limited to annotated exons

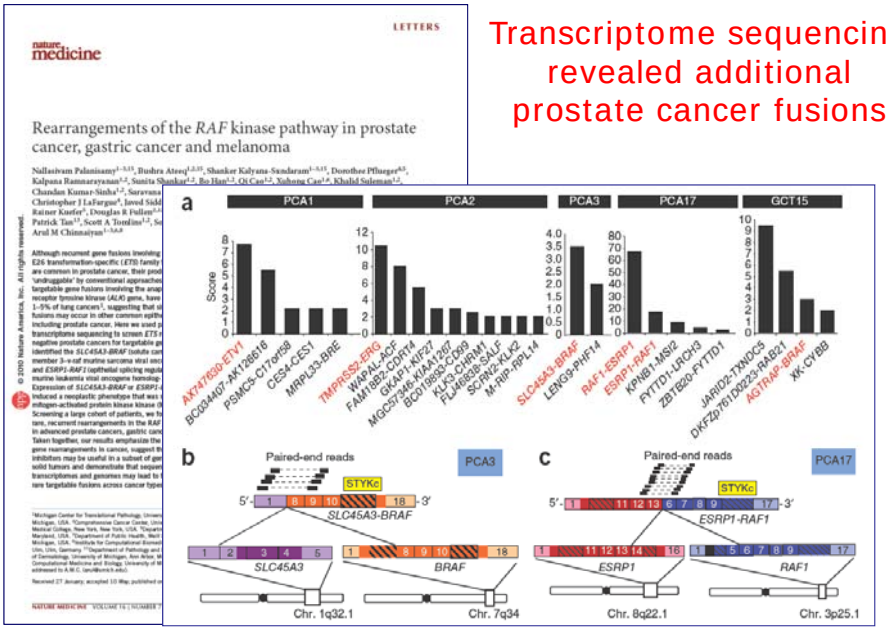
When biased is the goal

E.g. detection of mutations within

- a specific gene or genomic locus
  - long range PCR, capture with oligonucleotides
- all transcribed sequences
  - transcriptome sequencing
  - exome-capture



Transcriptome sequencing revealed additional prostate cancer fusions



Palanisamy *et al.*, Nat.Med., 2010

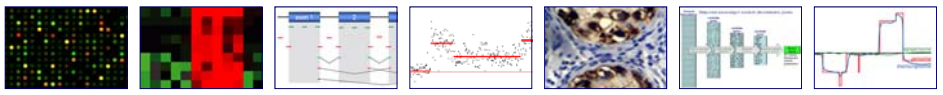
Genomic and personalised cancer biomedicine

Palanisamy *et al.*, Nat.Med., 2010

RAF-kinases, now also found involved in prostate cancers

Anti-RAF drugs already in clinical trials for melanoma; could be evaluated in lethal RAF-driven prostate cancers

➔ In the context of rational targeted therapy, cancers will be increasingly be classified by driving molecular events, rather than by organ site



Acknowledgements

The presented research data have been obtained in projects carried out at
Dept. Cancer Prevention, Inst. Cancer Res., Radium Hosp., Oslo University Hosp.

www.rr-research.no/cancerprevention