



“Next Generation Sequencing of Cancer Genomes”

Phil Stephens

WTSI :- Cancer Genome Project

Why is it important to study cancer?

Causes of cancer

How cancers develop

Pathways involved

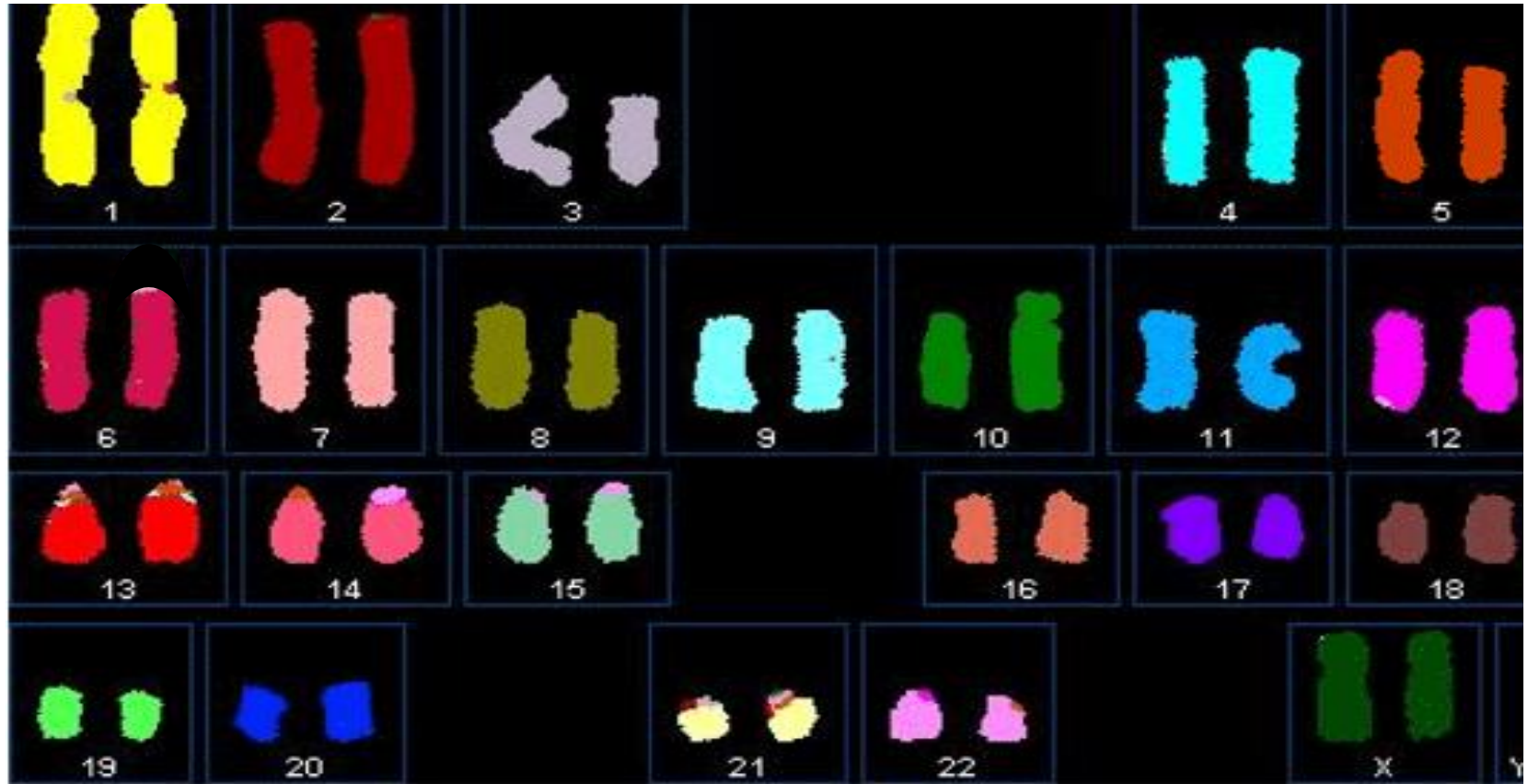
Development of tests for early detection

New targets for anticancer drug development

Monitor whether treatment is working

Challenges of studying cancer

46 Chromosomes

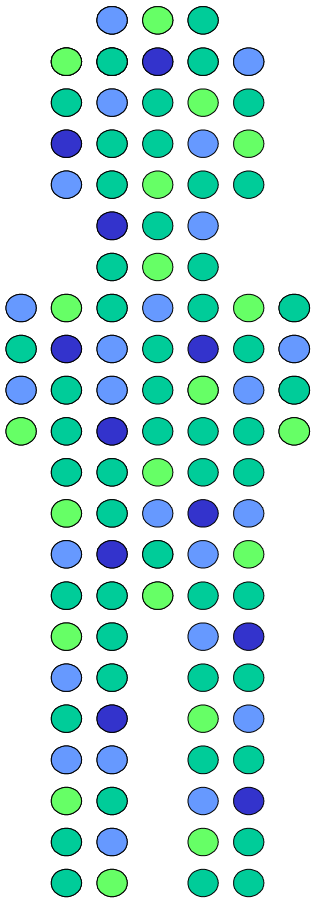


73 Chromosomes

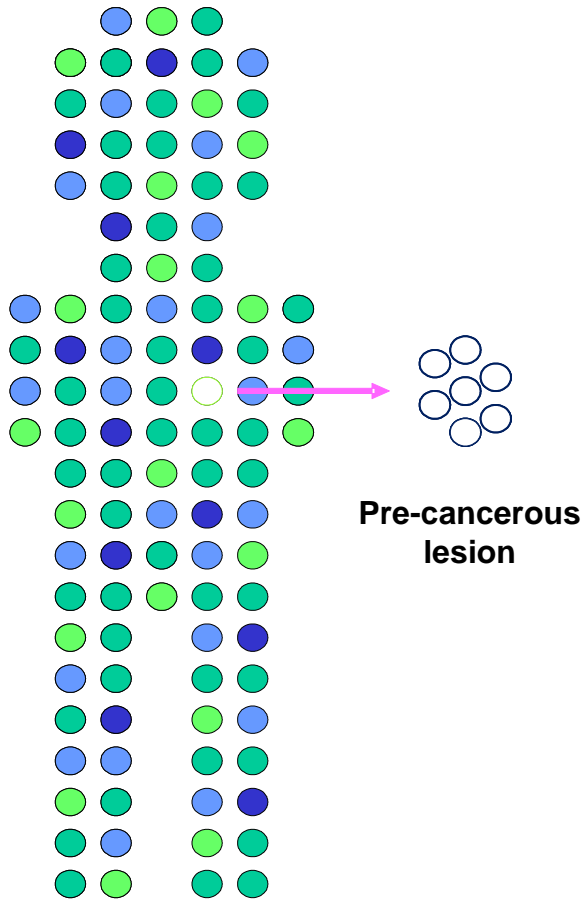


Multistep model of cancer

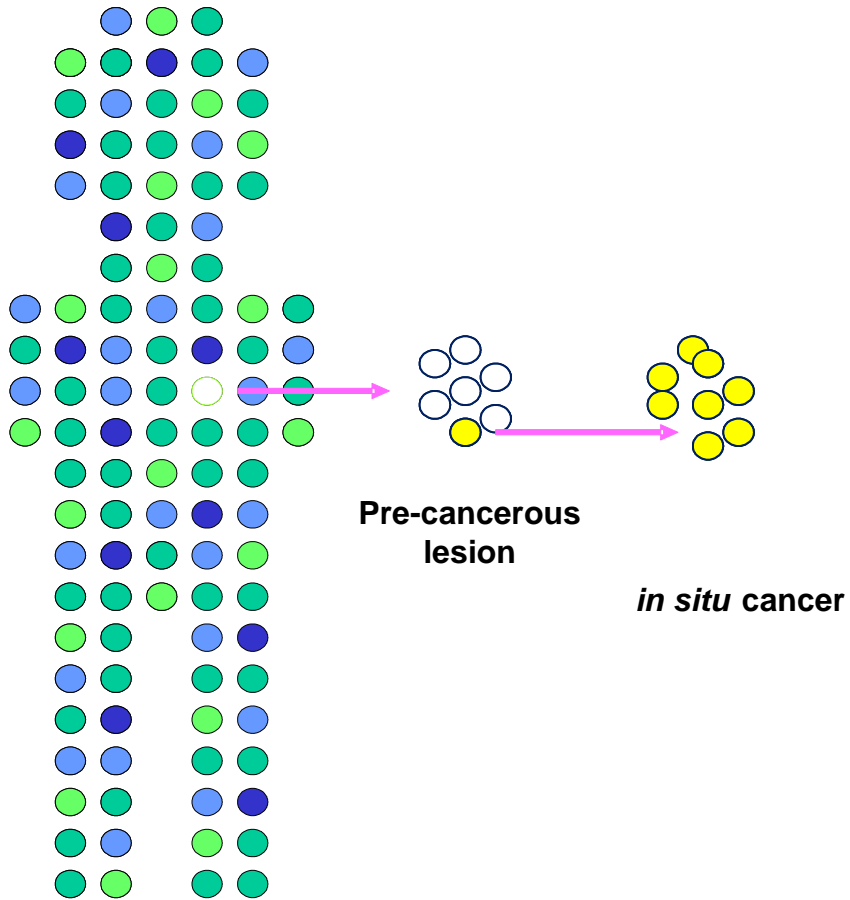
Multistep model of cancer



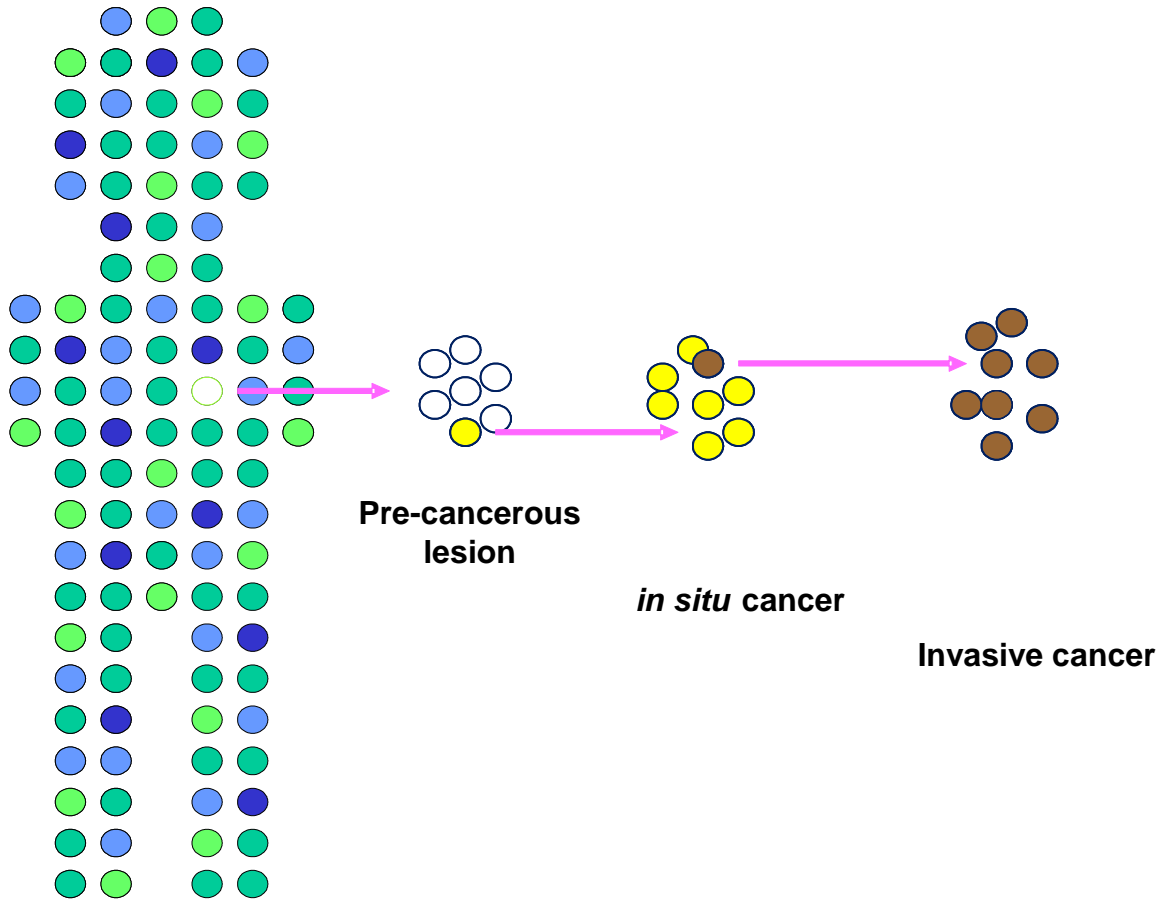
Multistep model of cancer



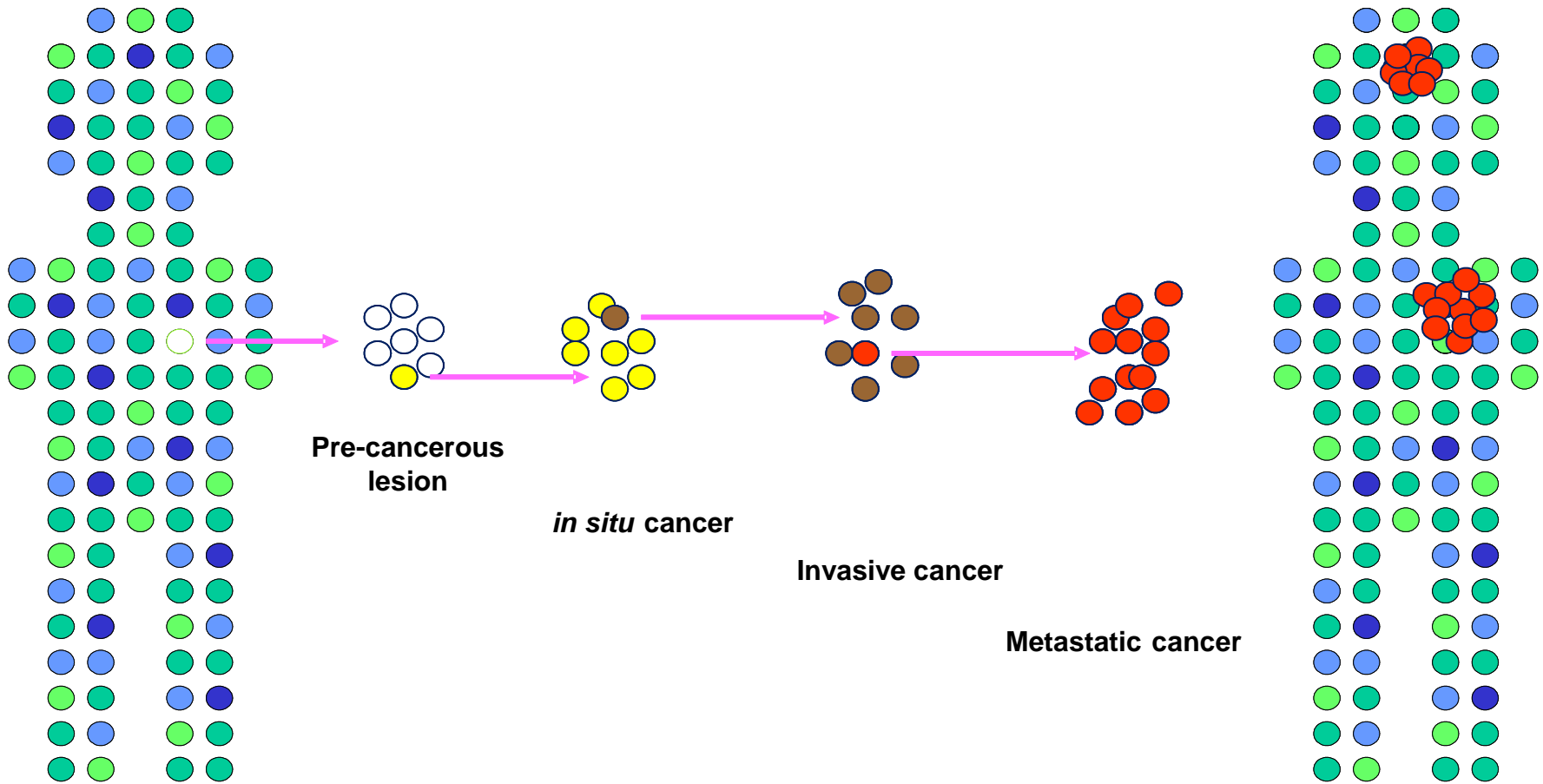
Multistep model of cancer



Multistep model of cancer

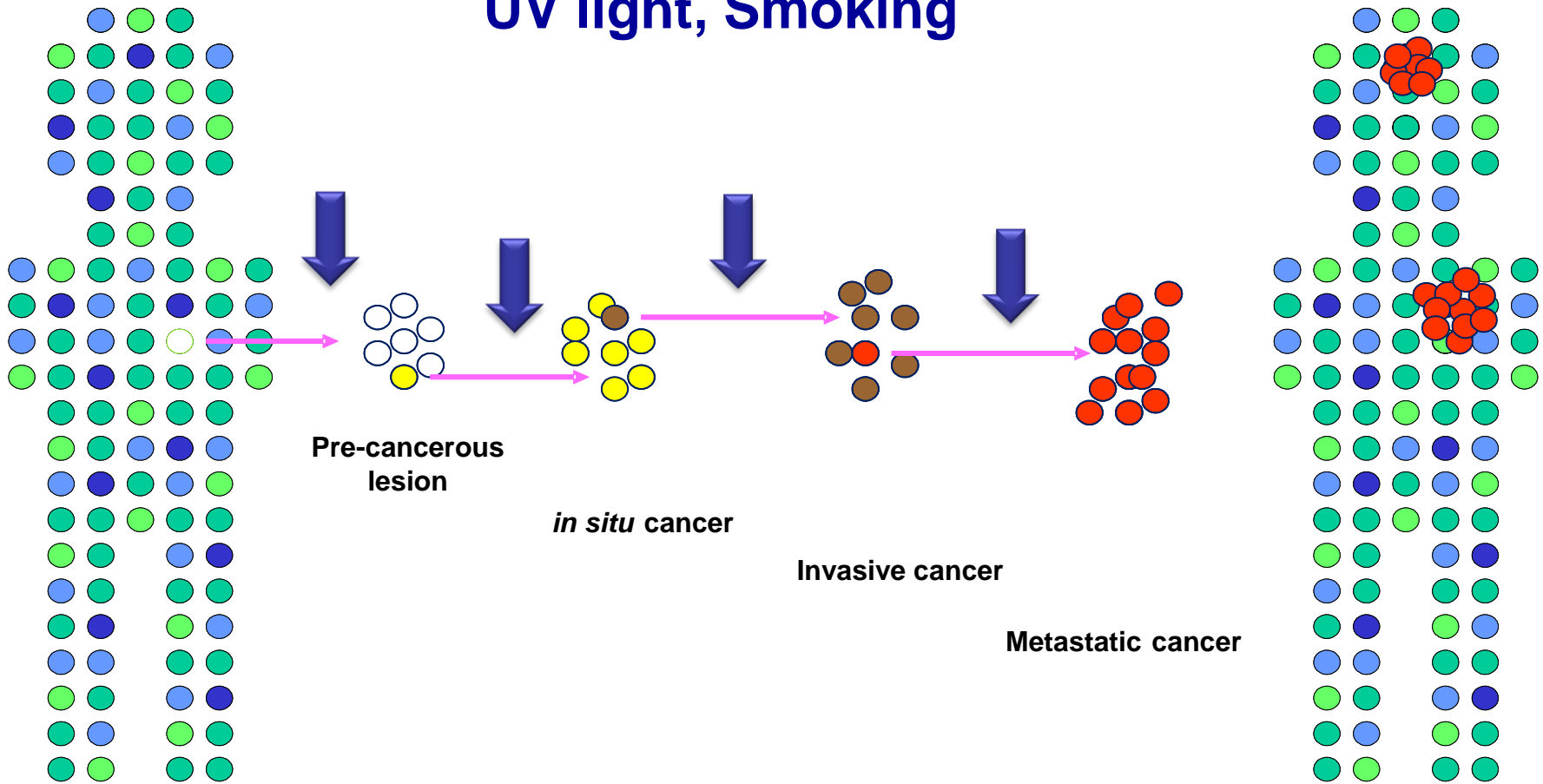


Multistep model of cancer

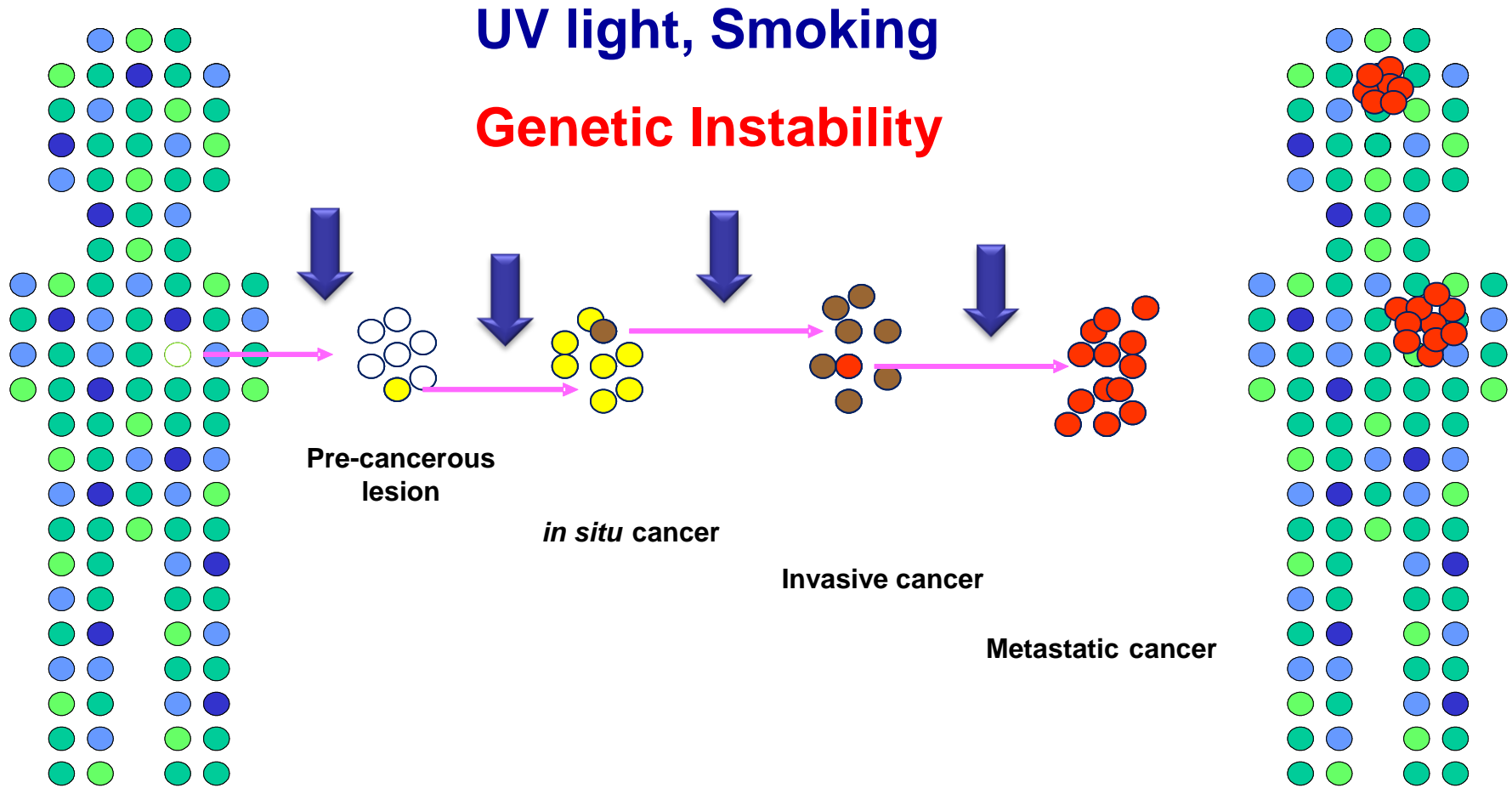


Development of complex cancer genomes

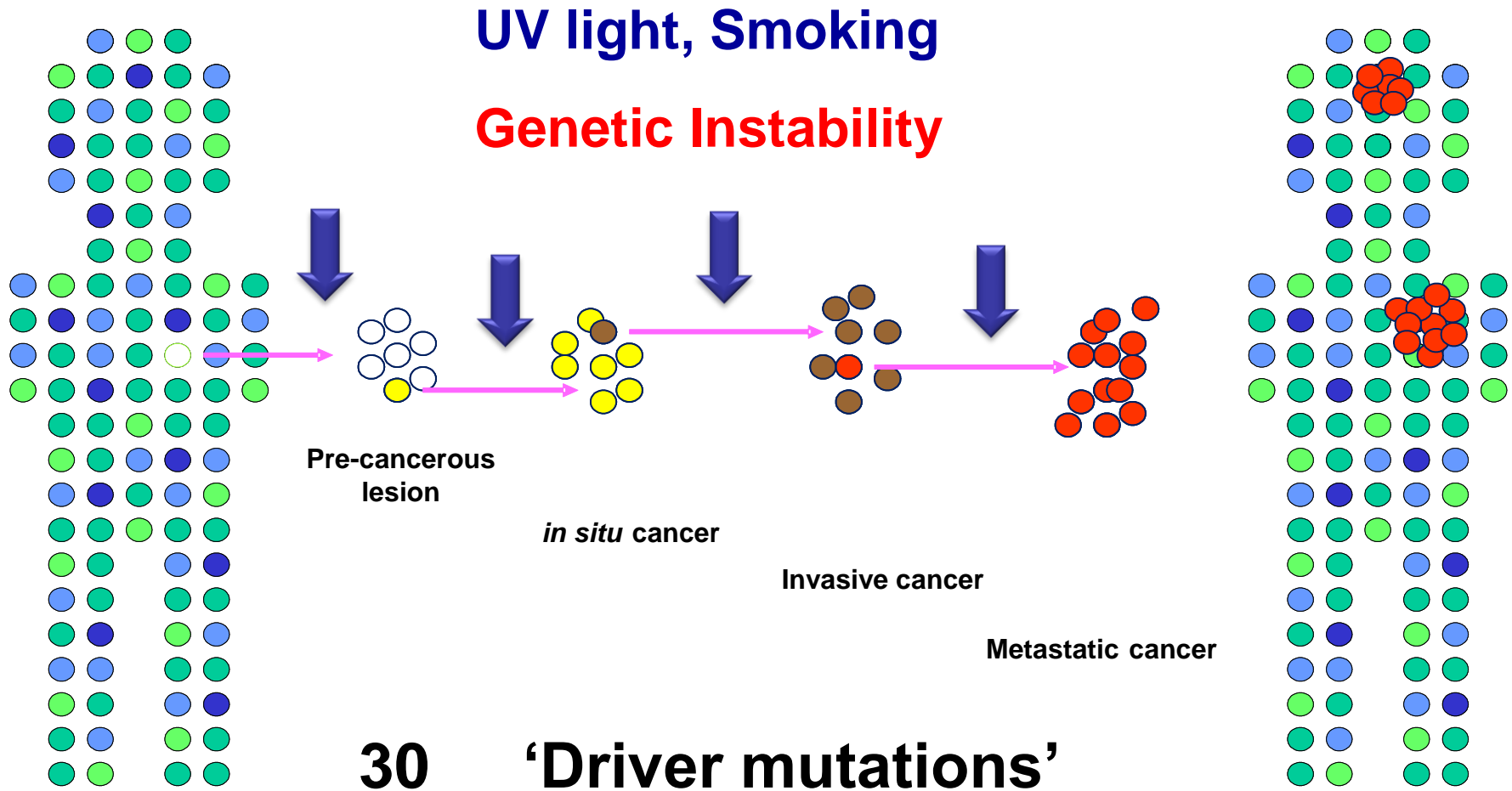
UV light, Smoking



Development of complex cancer genomes



Development of complex cancer genomes



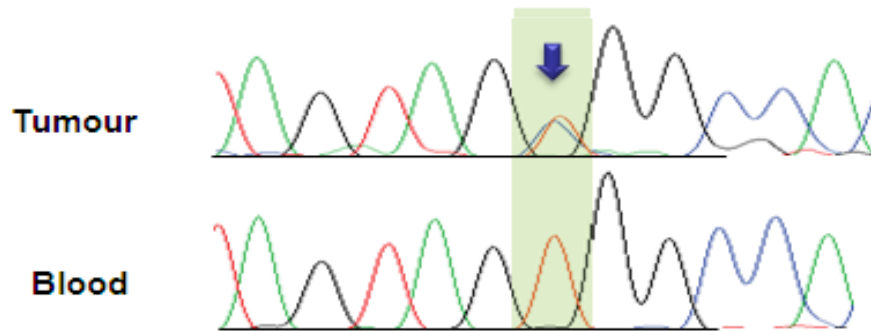
New sequencing technologies

100s of millions of DNA fragments simultaneously

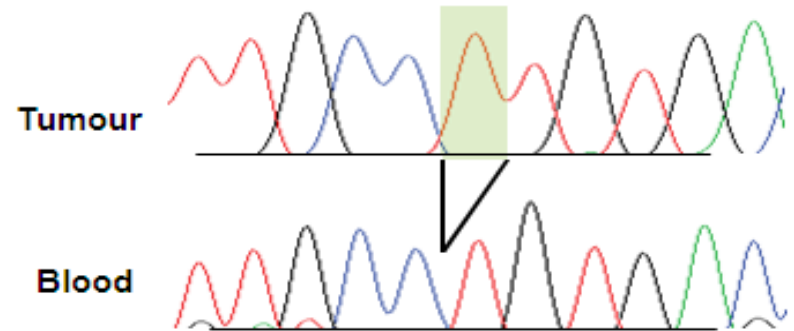
Essentially, 5 YEARS of data every DAY

Identify all classes of mutation in a single experiment

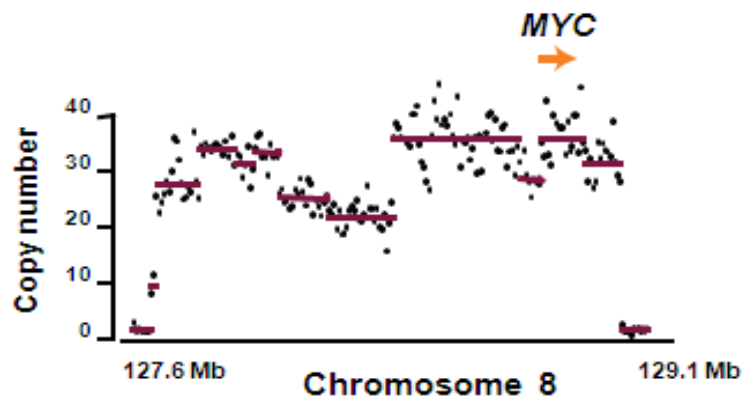
Base Substitutions



Insertions deletions



Copy number changes

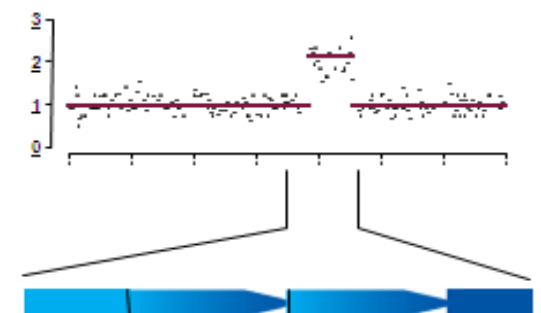


Rearrangements

Interchromosomal



Intrachromosomal



Next Generation Cancer Exome Sequencing

Investigate all protein coding exons/miRNA's simultaneously

Greater sensitivity to detect somatic variants

Can process large numbers of samples



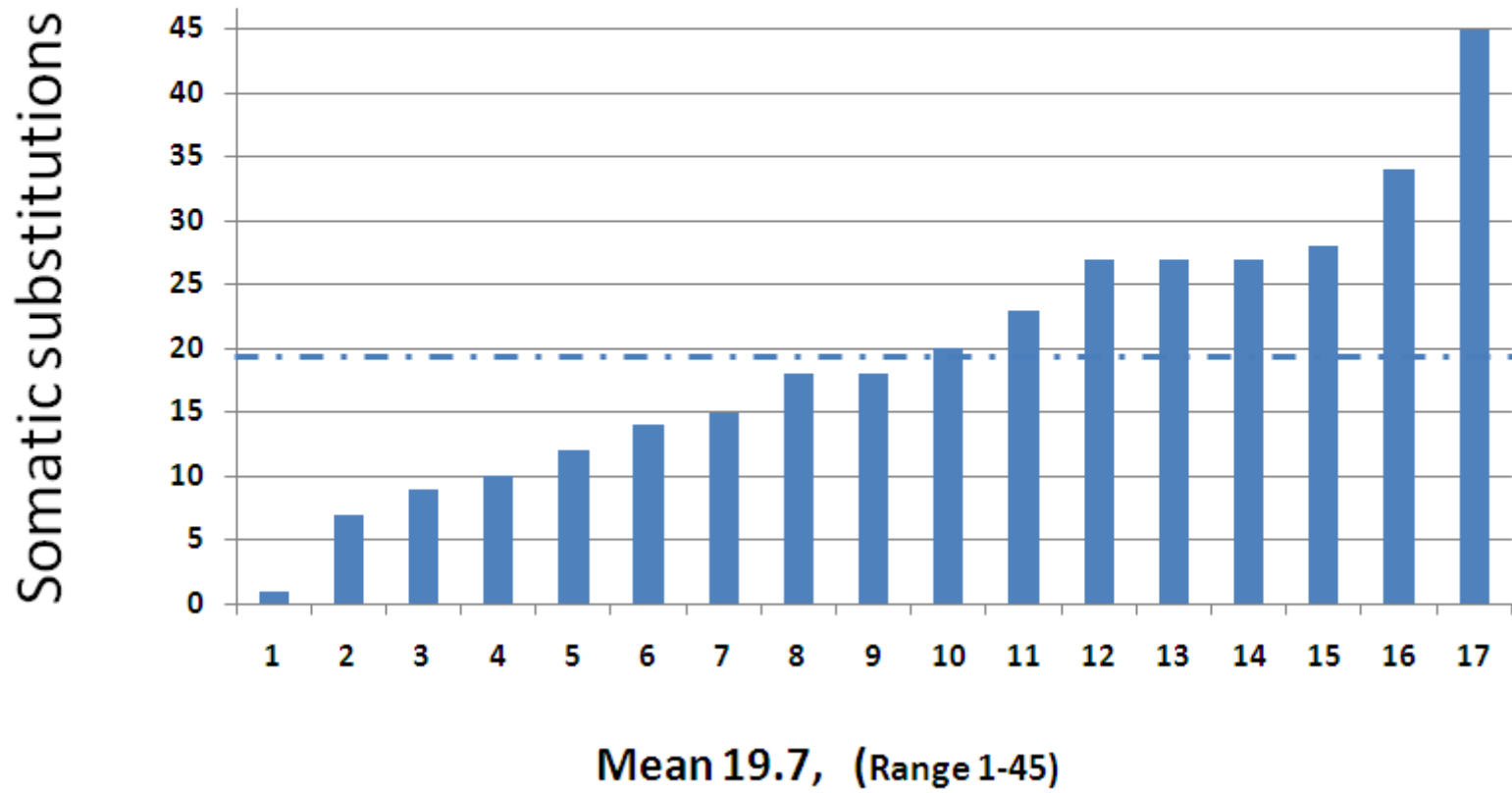
Breast cancer samples (28 total)

25 x ER⁺

3 x triple neg

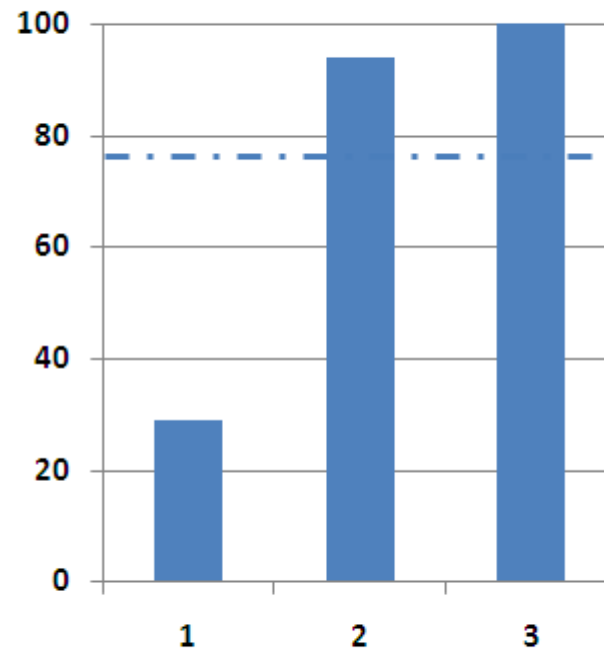
Distribution of coding somatic substitutions

ER⁺



Distribution of coding somatic substitutions

Triple neg



Mean 74.3, (Range 29-100)

Substitutions in known cancer genes

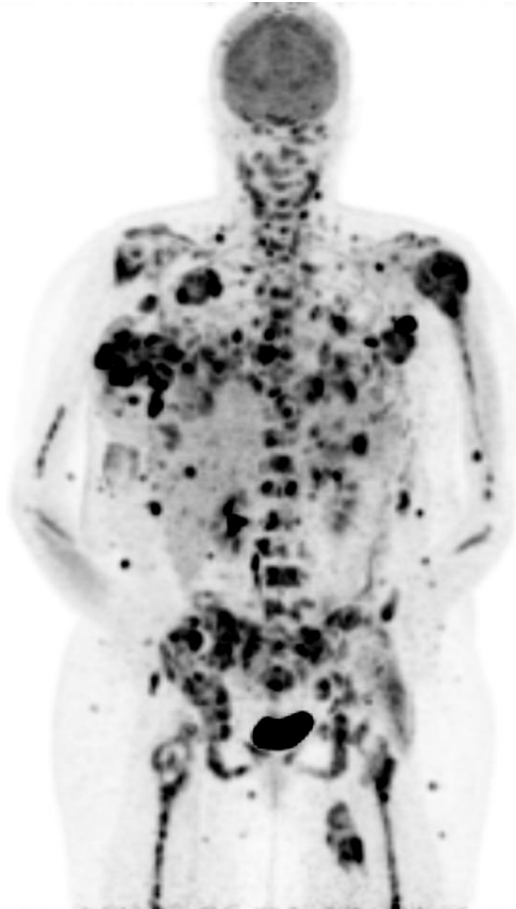
PD3995a	AKT1	E17K
PD3995a	NF1	G-1T
PD3994a	PIK3CA	N345K
PD3989a	PIK3CA	E545K
PD3856a	PIK3CA	H1047R
PD3857a	PIK3CA	H1047R
PD3888a	PIK3CA	H1047R
PD3983a	PIK3CA	H1047R
PD3985a	PIK3CA	H1047R
PD3992a	PIK3CA	H1047R
PD3996a	PTEN	Y27D
PD3991a	TP53	G245S
PD4002a	TP53	H179Y
PD3987a	TP53	Y220C
PD3986a	TP53	G+1A
PD3985a	TP53	R306X

Insertions & Deletions

Sample	Gene	Mutation
PD3849a	CDH1	V193X
PD3984a	MAP2K4	V151X
PD3992a	MAP2K4	I81X
PD3989a	PTEN	L370X
PD3995a	GATA3	N352X
PD3988a	GATA3	N352X
PD4004a	GATA3	Read through

Mutations in JNK and p38MAPK pathway
in >50% of ER+ breast cancer

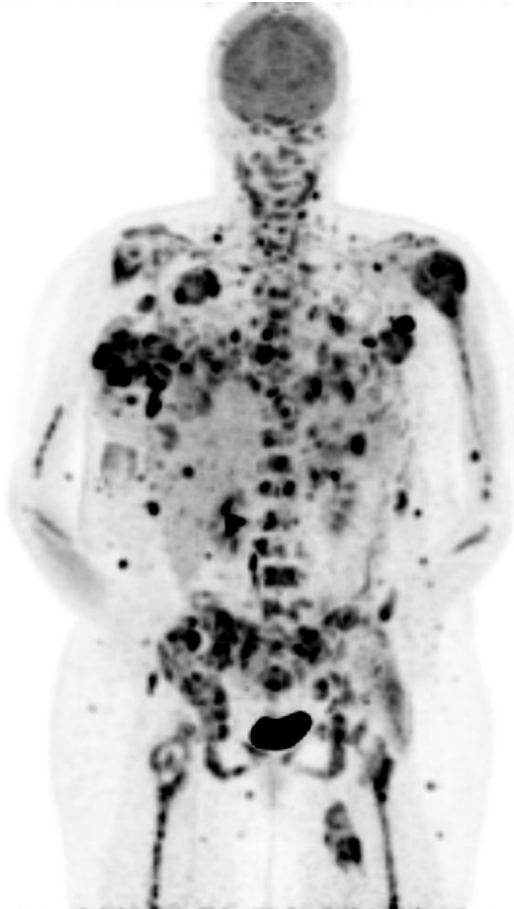
Patient with malignant melanoma



Courtesy of Dr Grant McArthur

Selective inhibitor of BRAF V600E

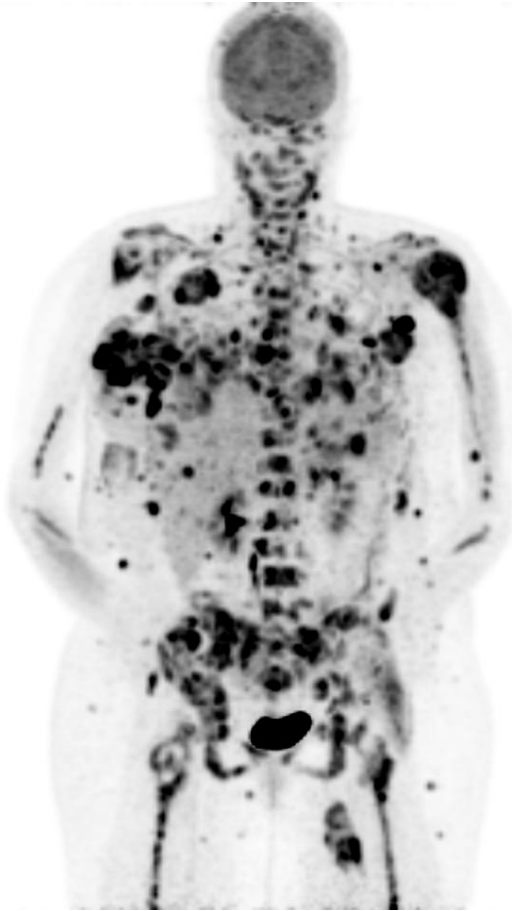
(Plexxicon 4032)



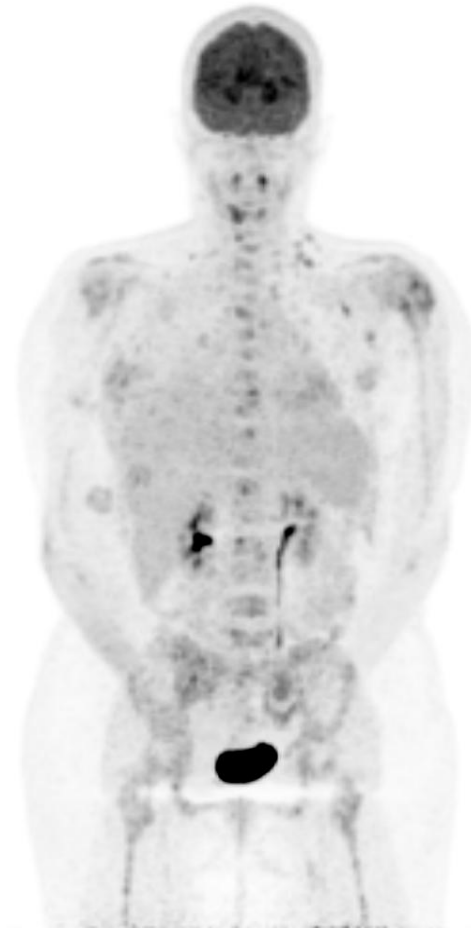
Courtesy of Dr Grant McArthur

Selective inhibitor of BRAF V600E

(Plexxicon 4032)



Before

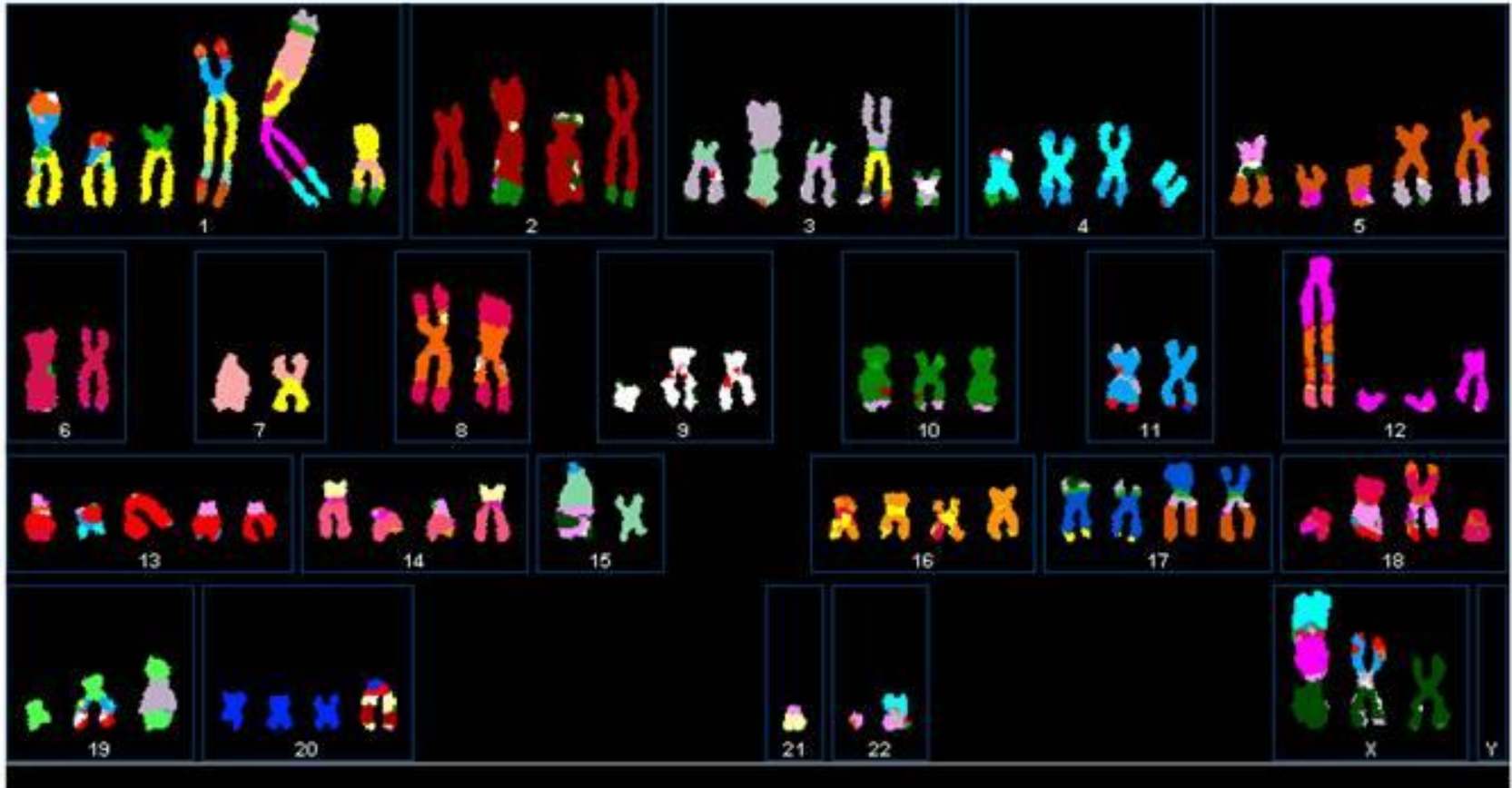


15 days after

Courtesy of Dr Grant McArthur

Rearrangement characterisation

Breast cancer HCC38 spectral karyotype



73 Chromosomes, 37 structural abnormalities

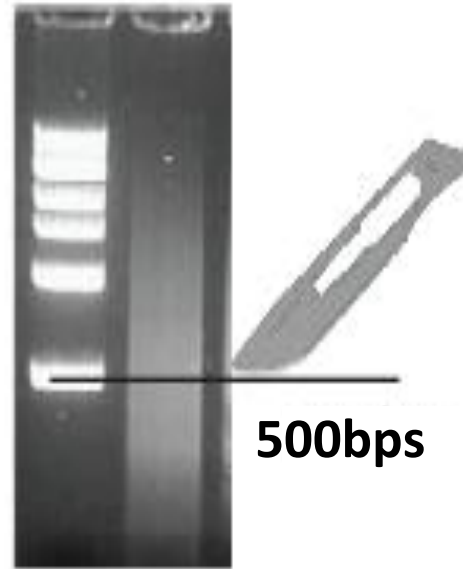
Summary of Illumina GA protocol

(i)



**Randomly shear DNA
& ligate Illumina adaptors**

(ii)



**Size select
& amplify**

(iii)



**Sequence from
both ends**

Summary of Illumina GA protocol



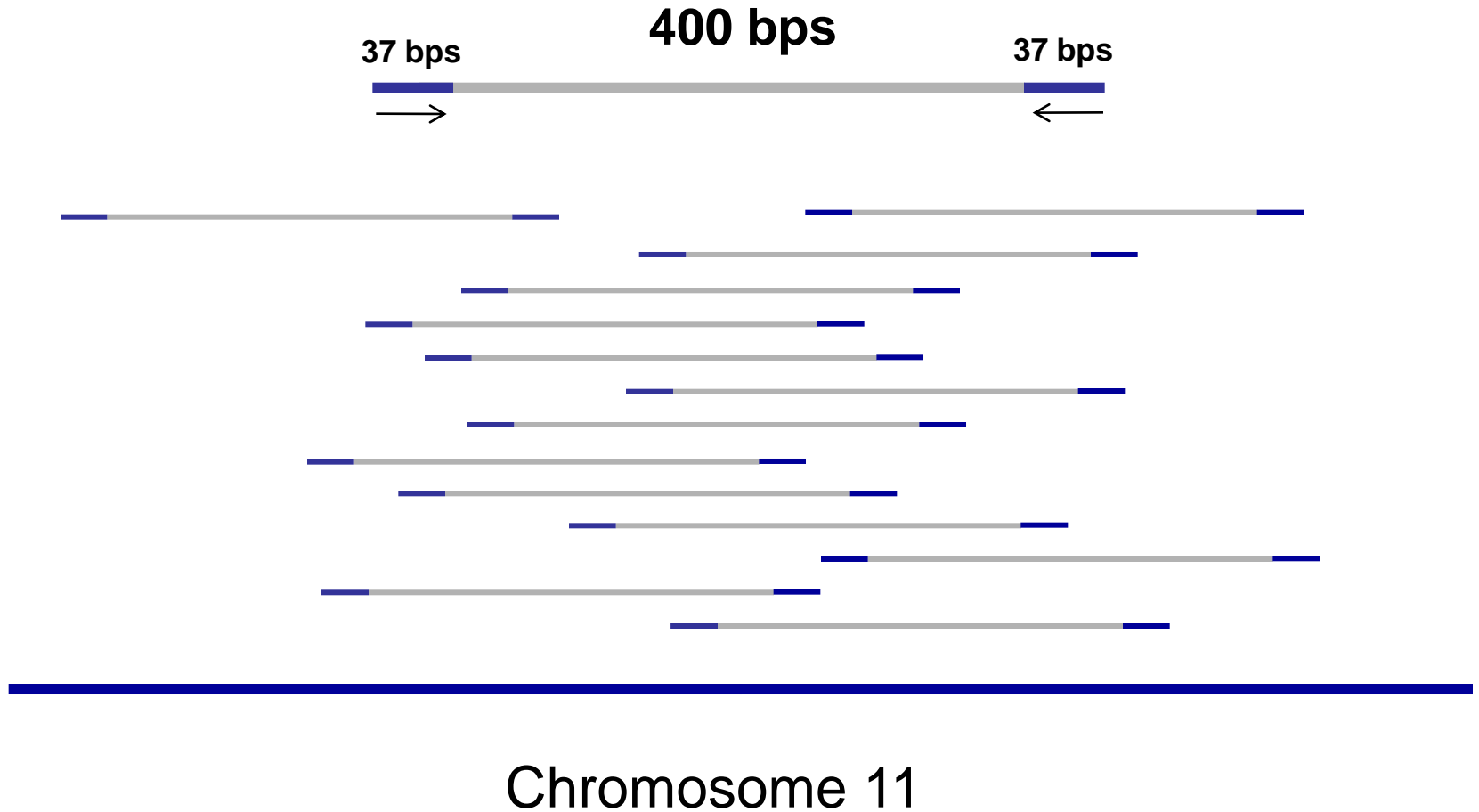
Summary of Illumina GA protocol



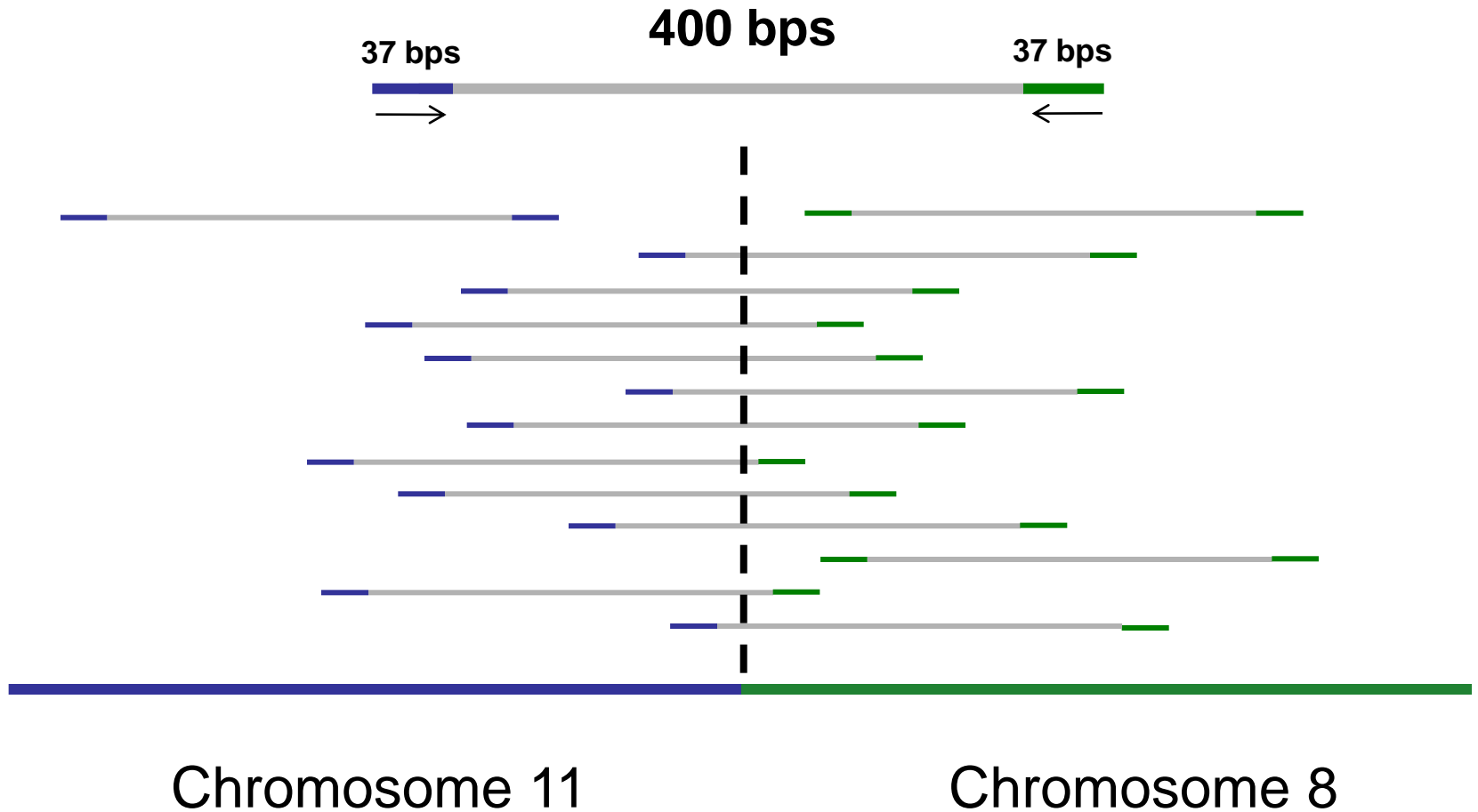
Align reads to ref sequence

MAQ algorithm

Correctly mapping paired end reads



Incorrectly mapping paired-end reads

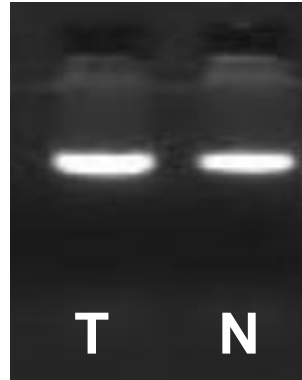


PCR amplify in tumour and matched normal

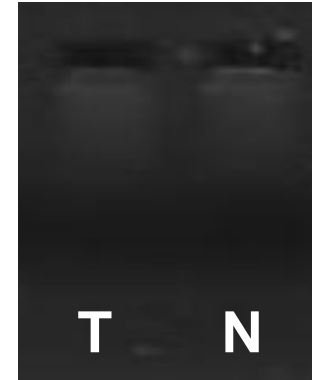
Somatic



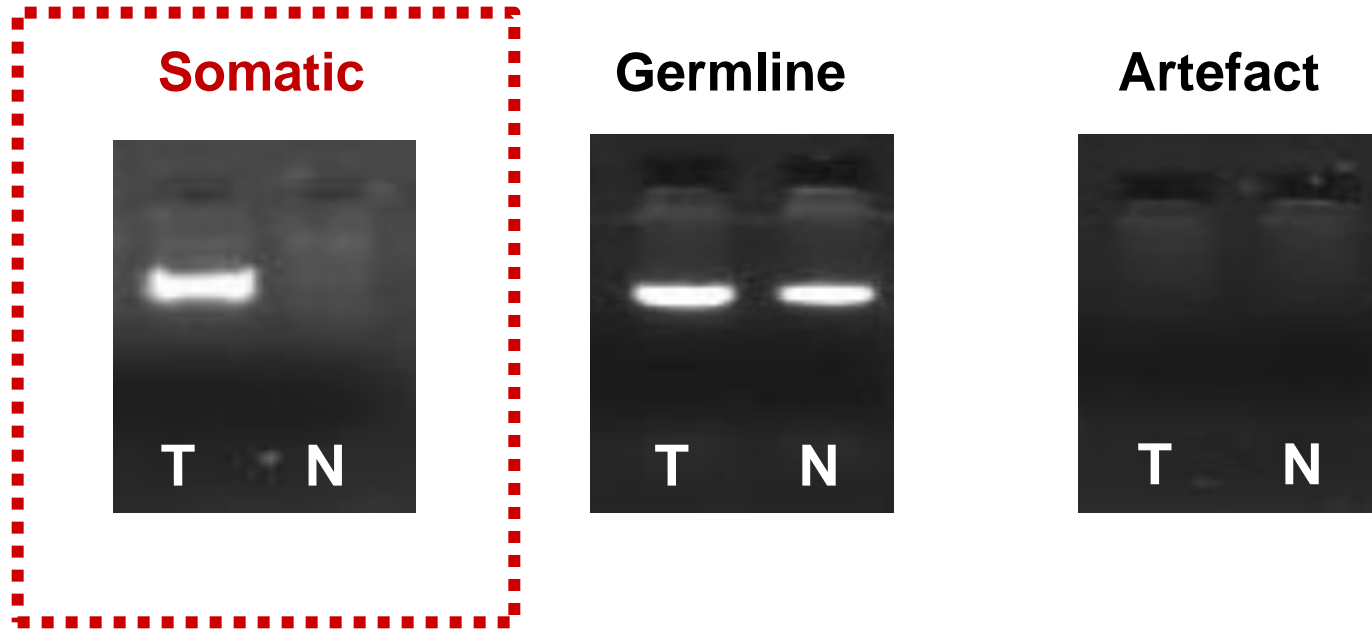
Germline



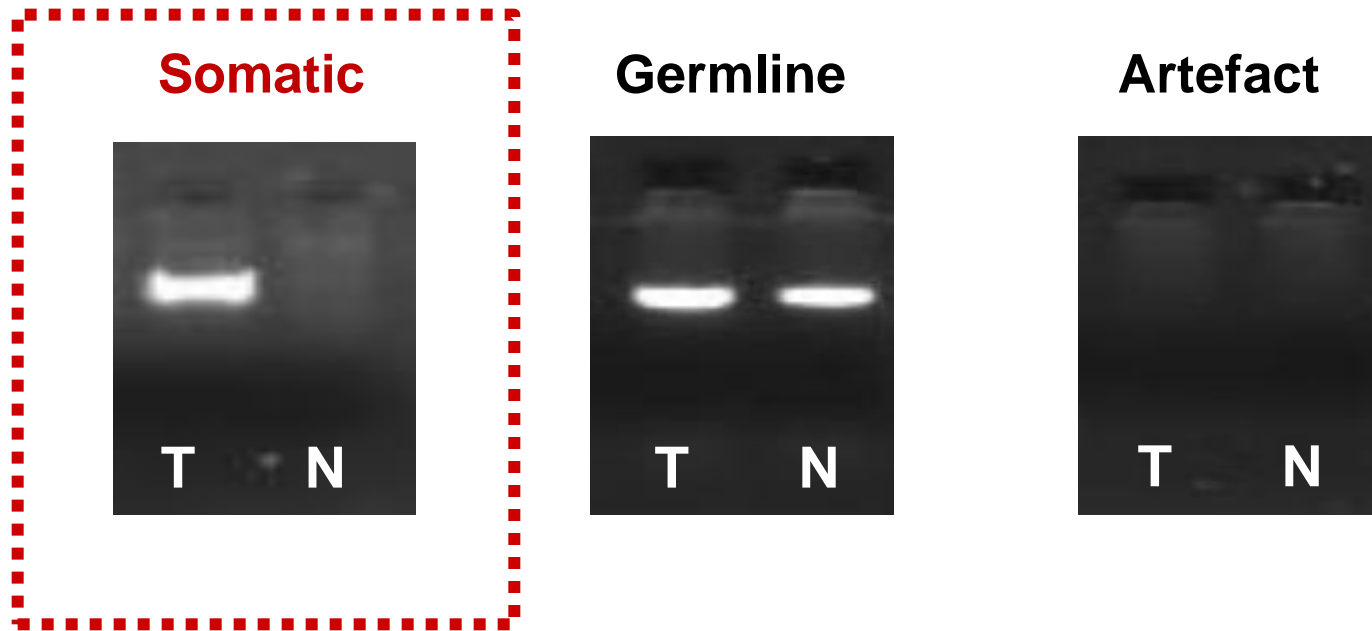
Artefact



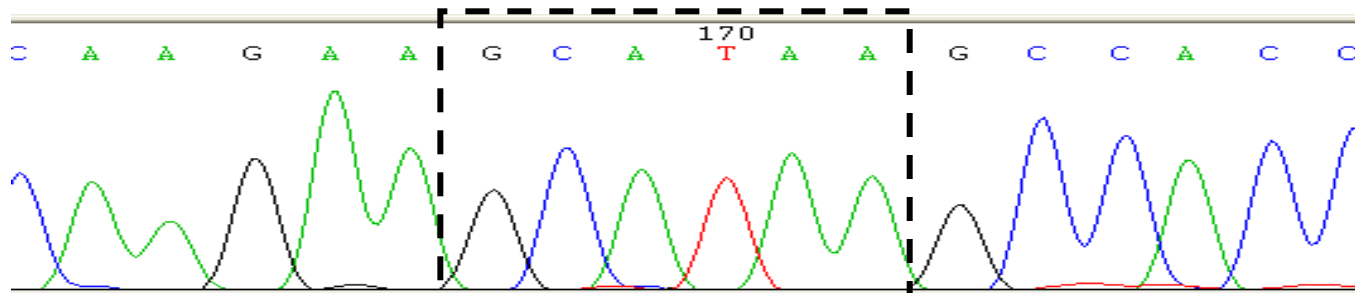
PCR amplify in tumour and matched normal



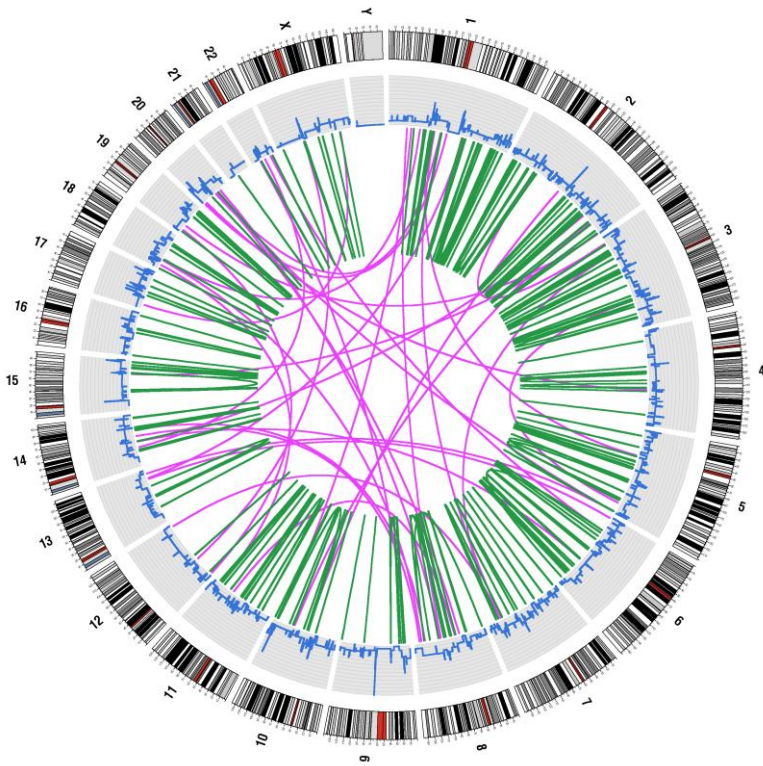
PCR amplify in tumour and matched normal



6 bps of microhomology at break

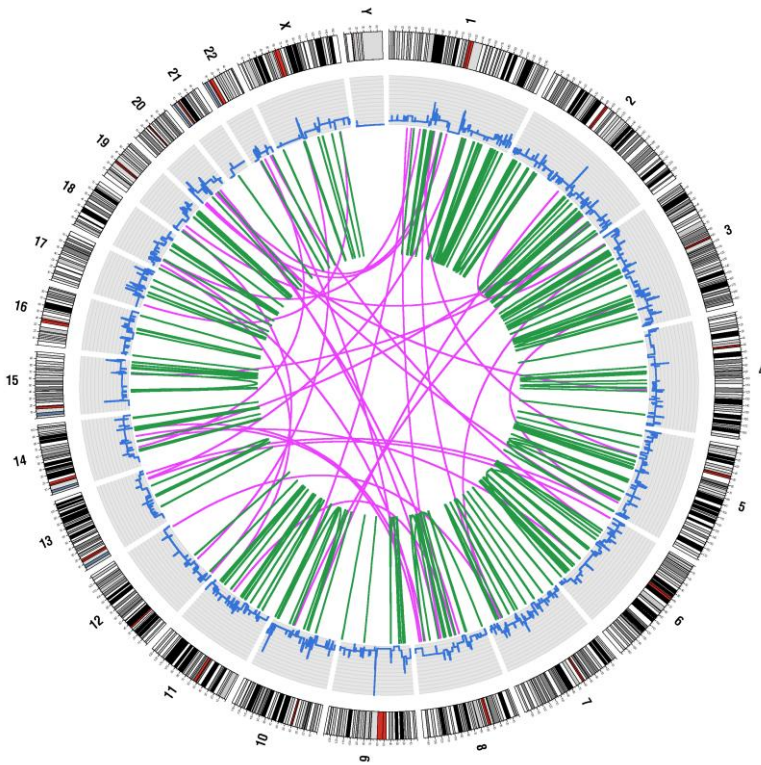


Breast cancer HCC38 (Triple neg)

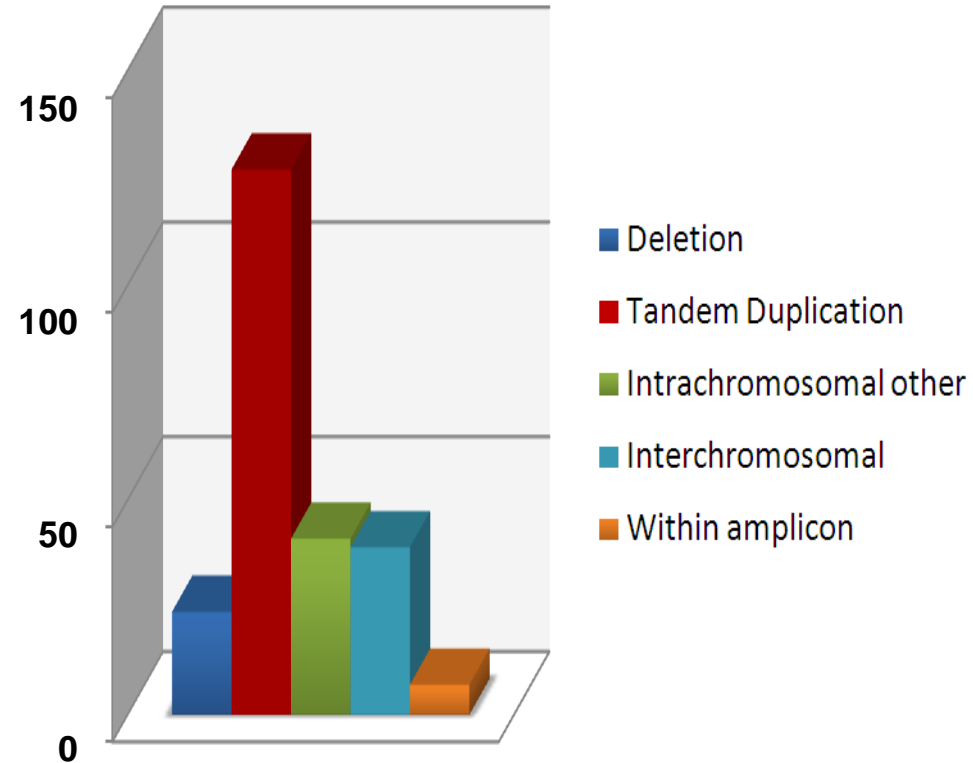


238 somatic structural variants

Breast cancer HCC38 (Triple neg)



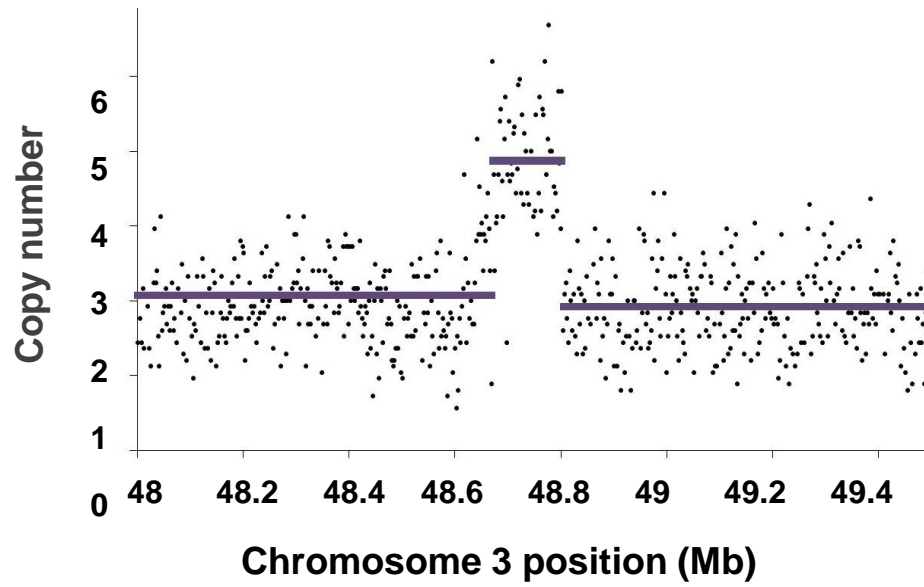
238 somatic structural variants



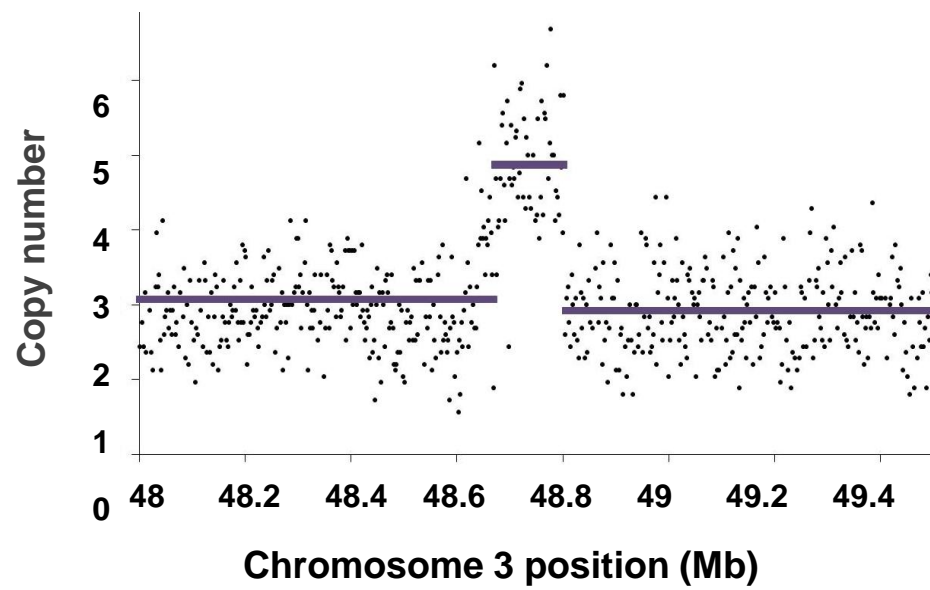
Patterns of variation

What are these structural variants doing?

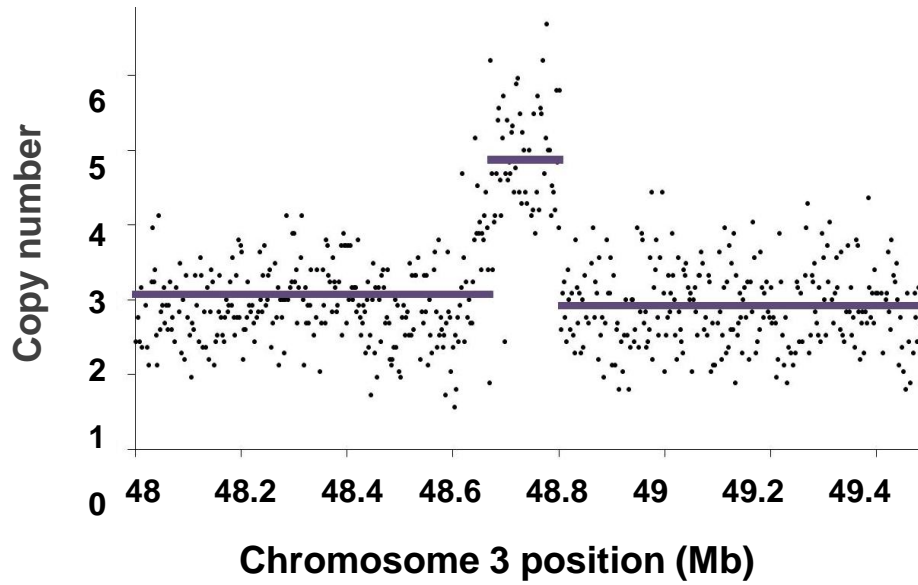
~160 kb copy number increase



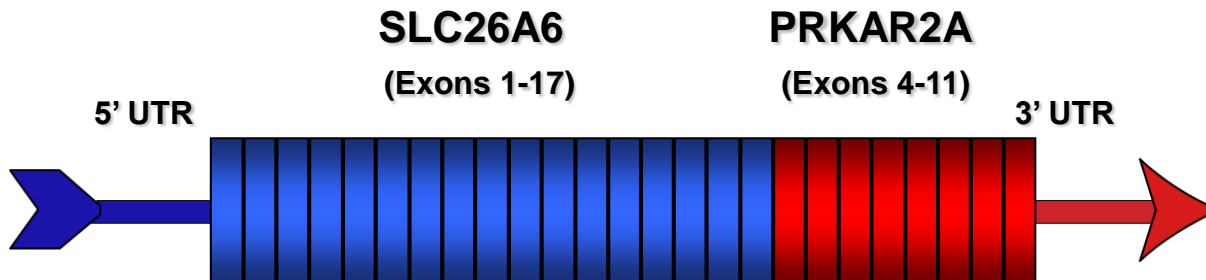
164 kb tandem duplication



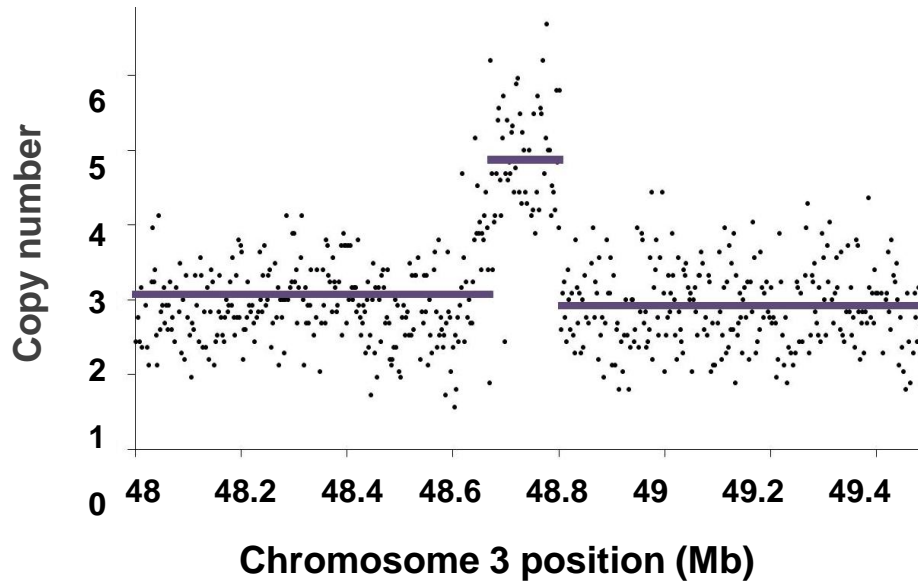
164 kb tandem duplication



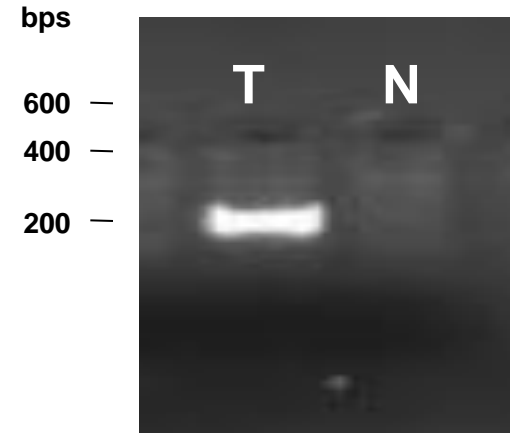
SLC26A6/PRKAR2A in frame fusion gene



164 kb tandem duplication



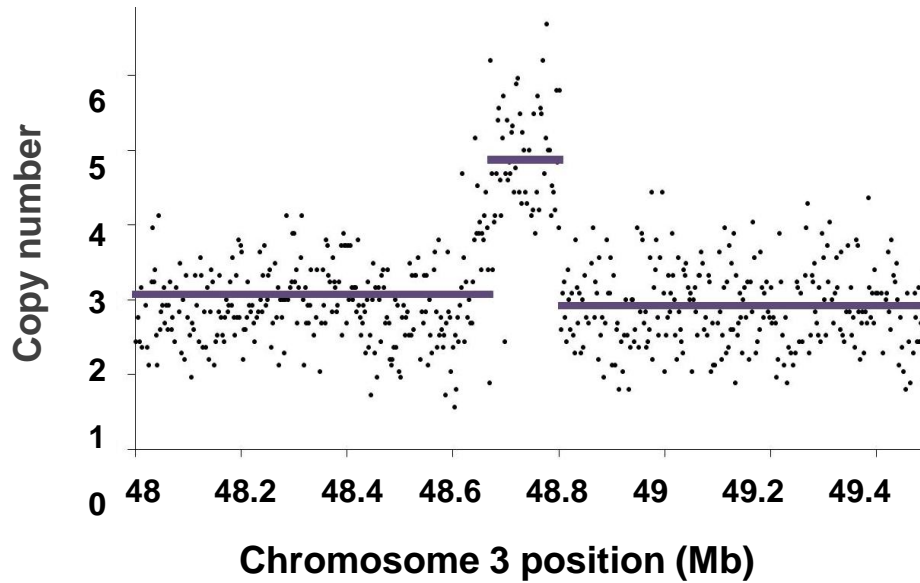
Genomic PCR



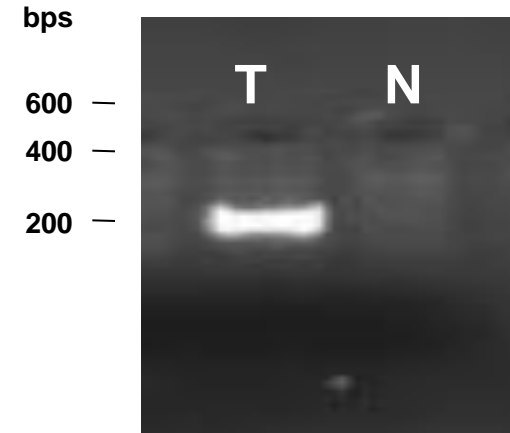
SLC26A6/PRKAR2A in frame fusion gene



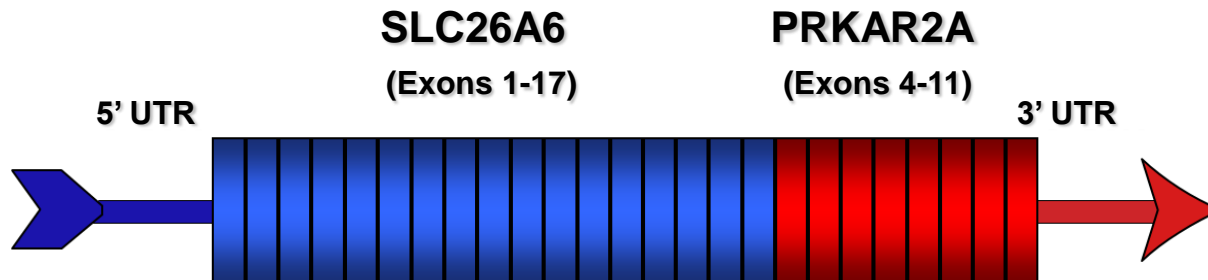
164 kb tandem duplication



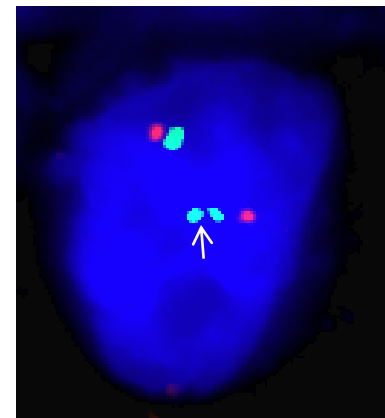
Genomic PCR

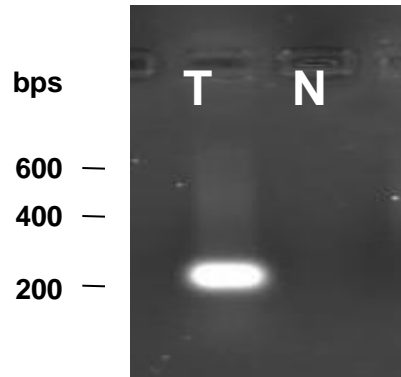


SLC26A6/PRKAR2A in frame fusion gene

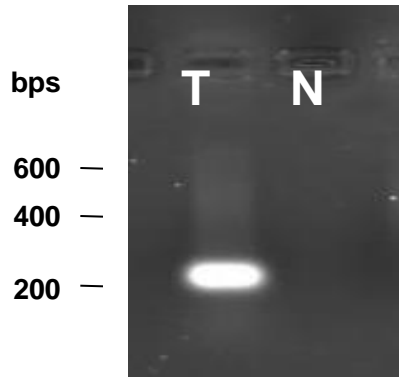


FISH confirmation of tandem duplication

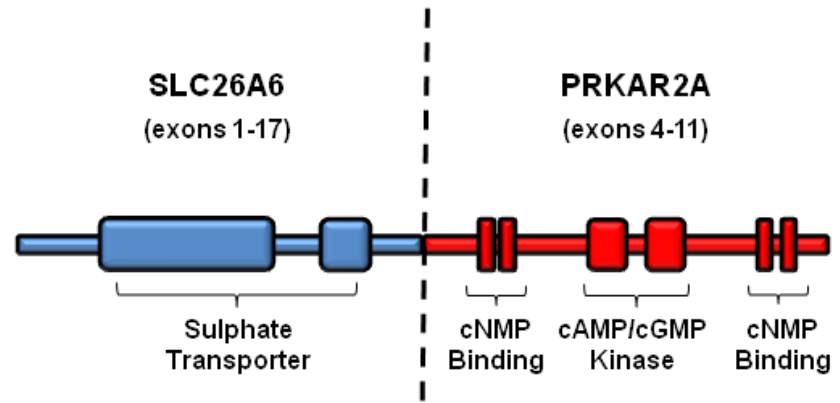


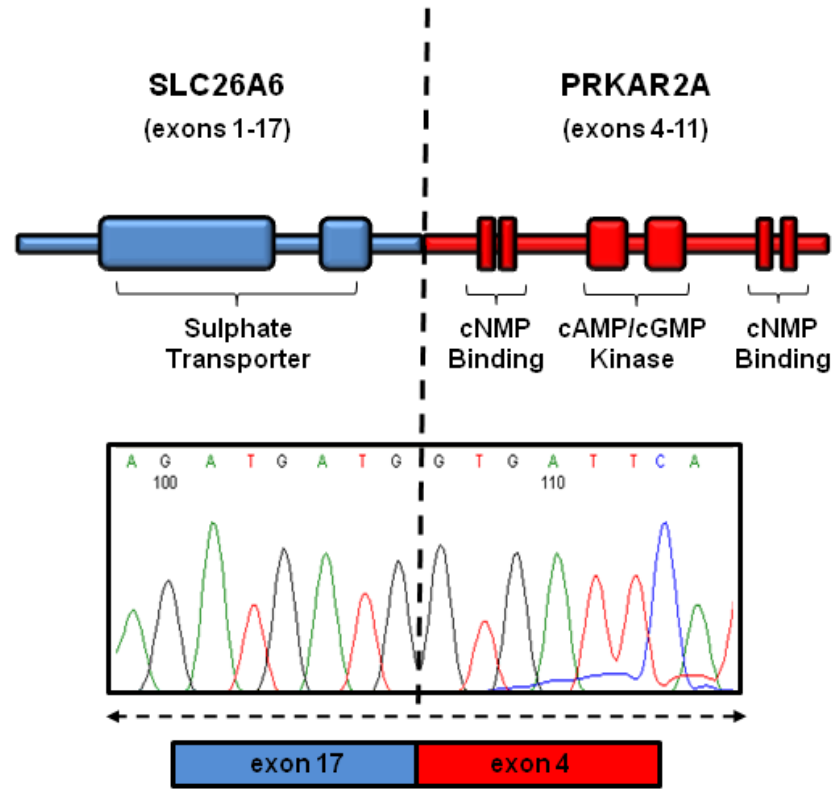
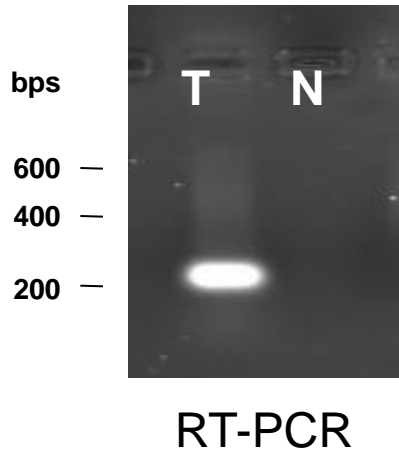


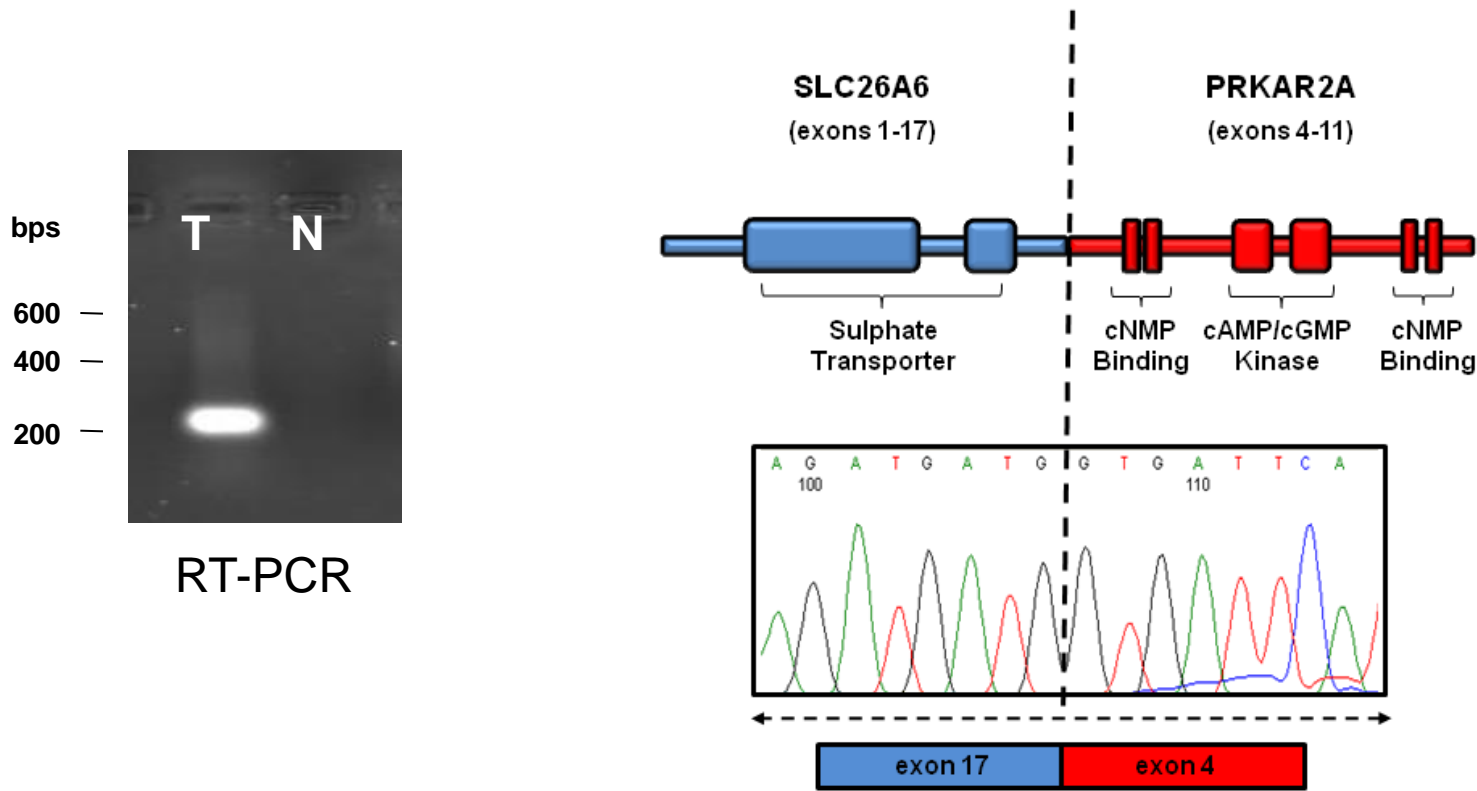
RT-PCR



RT-PCR







Predicted 914 amino acid fusion protein

MGLADASGPRDTQALLSATQAMDLRRRDYHMERPLLNQEHLEELGRWGSAPRTHQWRTWLQCSRARAYALLLQHLPVLVWLP RYPVVRDWLLGDLLSGL
 SVAIMQLPQGLAYALLAGLPVFLYSSFYVPFIYFLFGTSRHISVGTFAVMSVMVGSVTE LAPQALNDSMINETARDAARVQVASTLSVLVGLFQVGLGLI
 HFGFVVTYLSEPLVRGYTTAAAVQVFVSQLKYVFLHLSHSGPLSLIYTVLEVCWKLPQSKVGTVVTA AVAGVLLVVVKKLNDK LQQQLPMPPIGELLTIGAT
 GISYGMGLKHRFEVDVVGNI PAGLVPPVAPNTQLFSKLVGSAFTIAVVGFAIAISLGKIFALRHG YRVD SNQELVALGLSNLIGGIFQCFVSCSMSRSLVQEST
 GGNSQVAGAISSLFILLIIVKLGELFHDLPKAVLAAIIIVNLKGMLRQLSDMRSLWKANRADLLIWLVTFTATILLNLDLGLVVAVIFSLLLVVVRTQMPHYSVLGQ
 VPDTDIYRDVAEYSEAKEVRGVKVRSSATVYFANA EFYSDALKQRCGV DVF LISQKKLLKKQEQLK LKQLQKEEKL RQQAASPKGASVSIN VNTSLED MR
 SNNVEDCKMVIHPKTDEQRCRLQEACKDILLFKNL DQEQLSQVLDAMFERIVKADEHVIDQGD DGNFYVIERGT YDILVT KDNQTRSVGQYDNRG SFGEL
 ALMYNTPRAATIVATSEGLWGLDRVTFRRRIIVKNN AKKRKM FESFIESVPLLSLEVSE RMIKIVDVIG EKIKYK DGERIITQGEKADSFYIIESGEVSILIR SRTKSN
 KDGGNQEVEIARCHKGQYFGELALVTNK PRAASAYAVGDVKCLVMDVQAFERLLGPCMDIMKRNISHYEEQLVKMFGSSVDLGNLGQStop

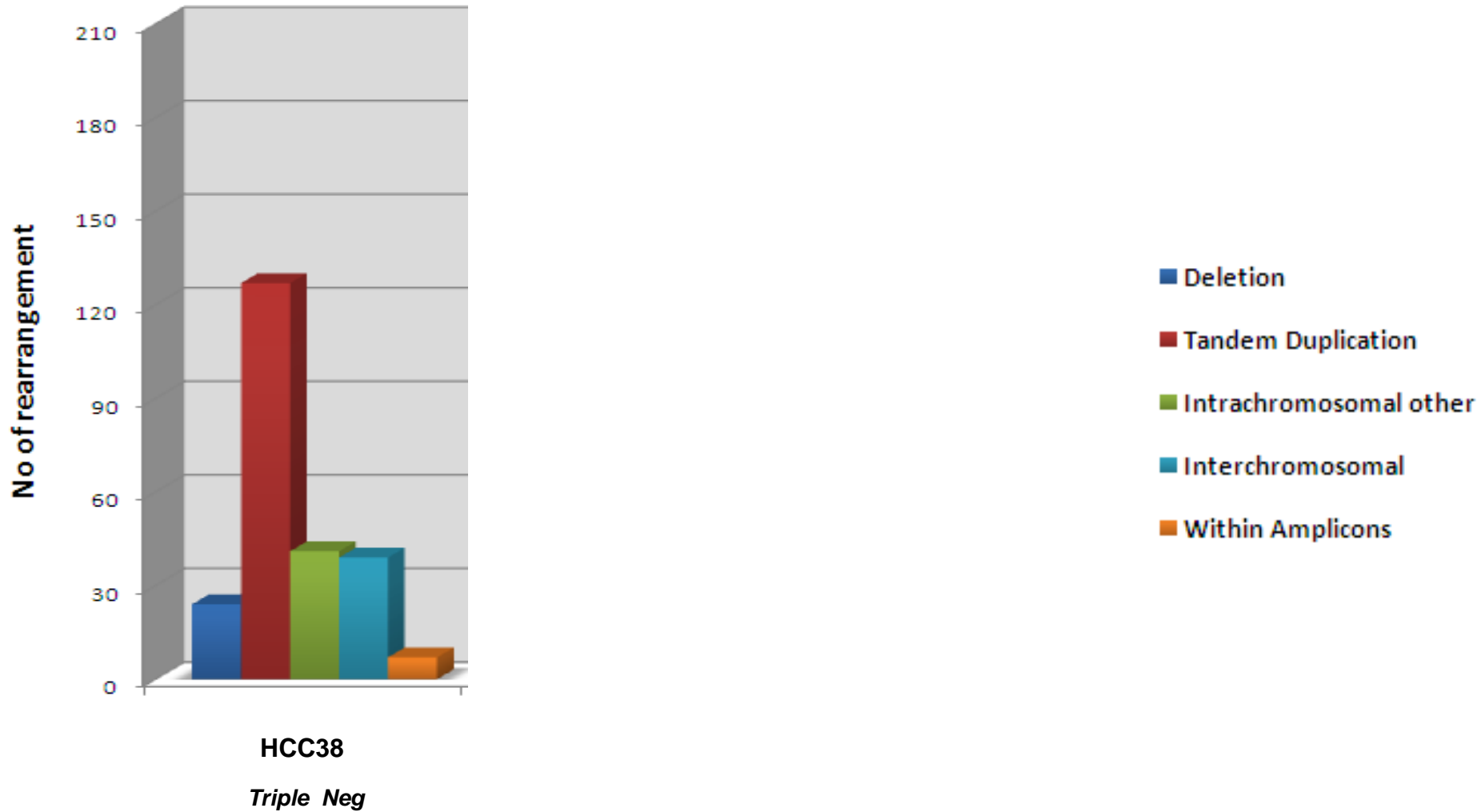
Five expressed in frame fusion genes

2 generated by tandem duplications

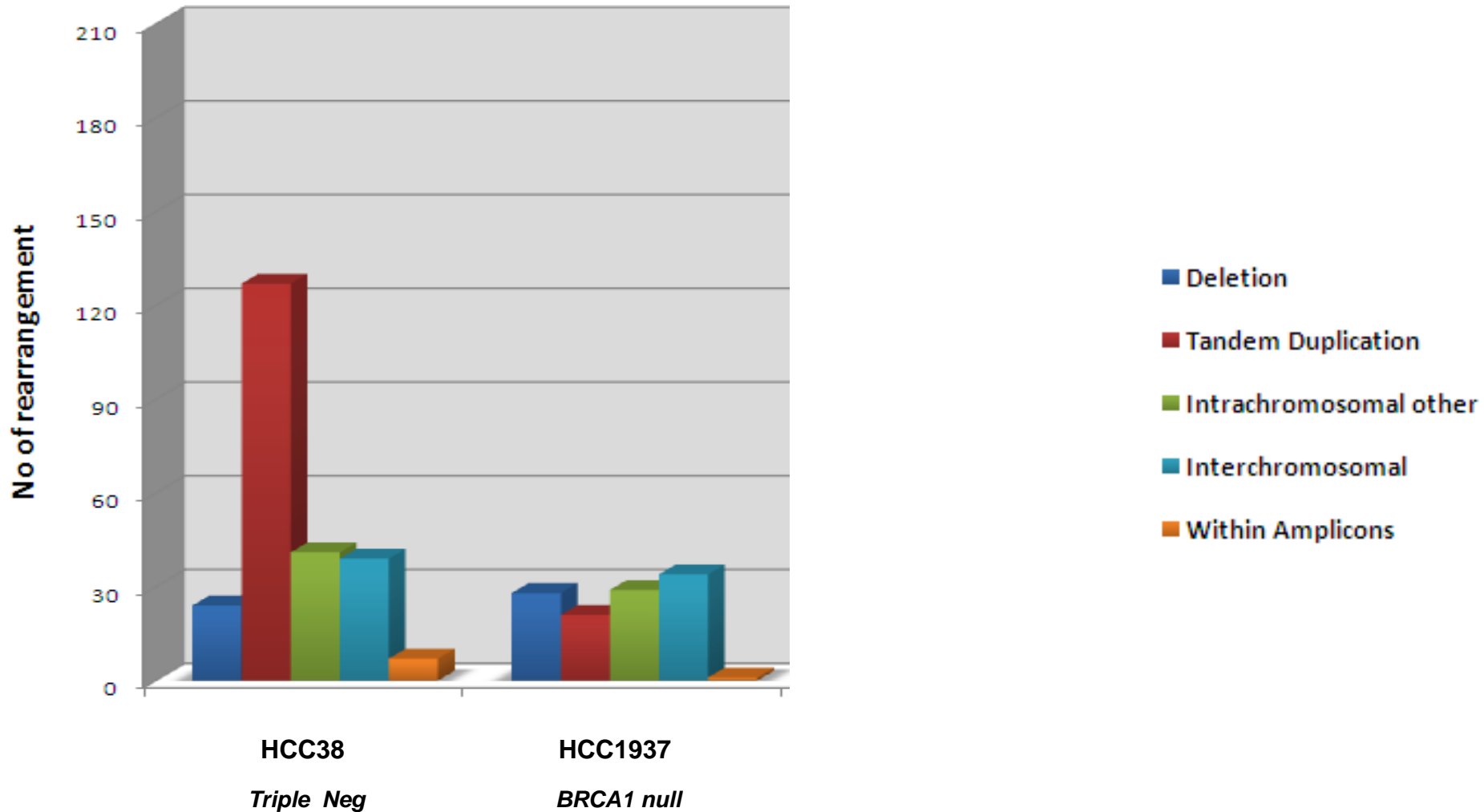
3 generated by large inversions

Different patterns of structural variation
are emerging from other breast cancers

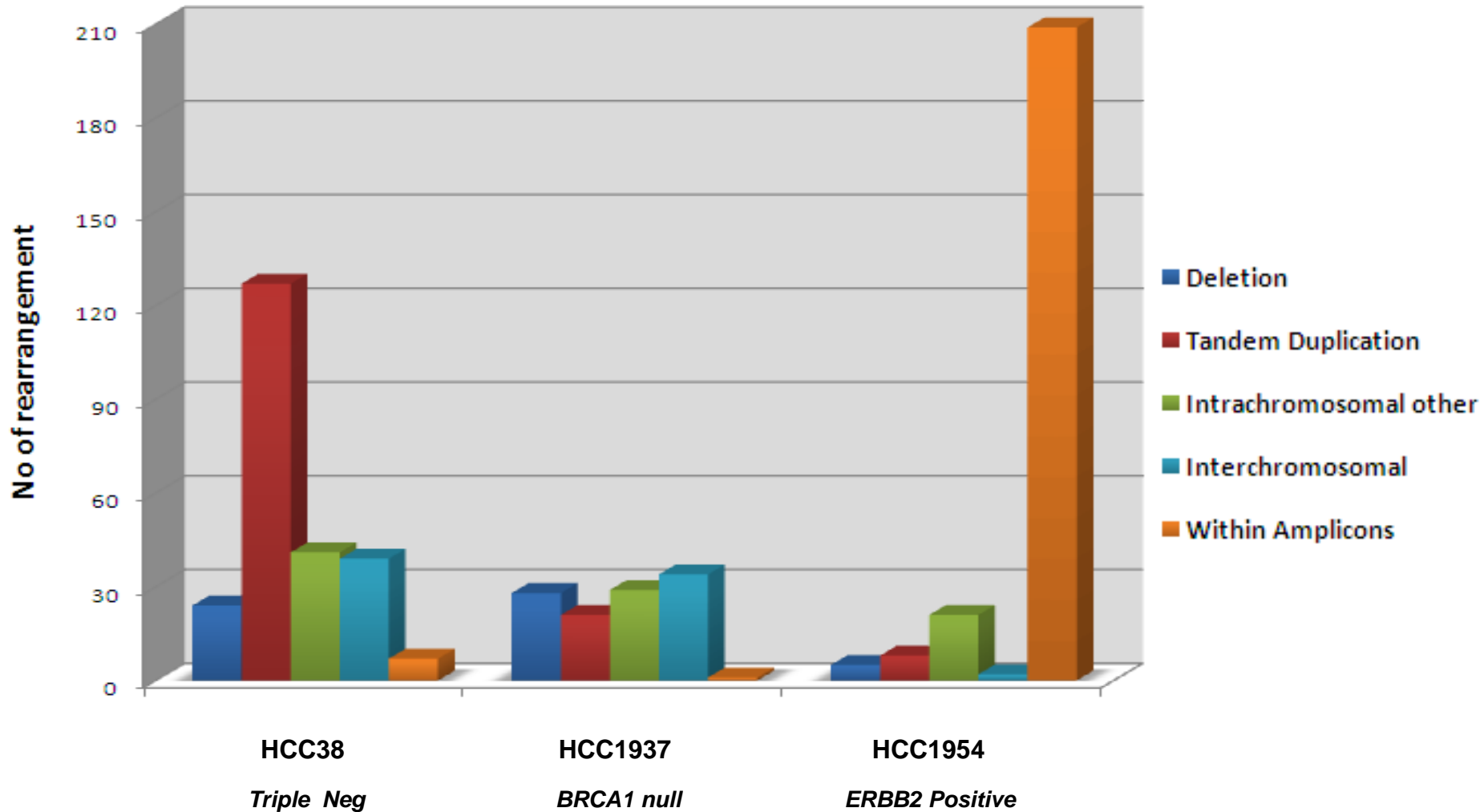
Patterns of somatic structural variation



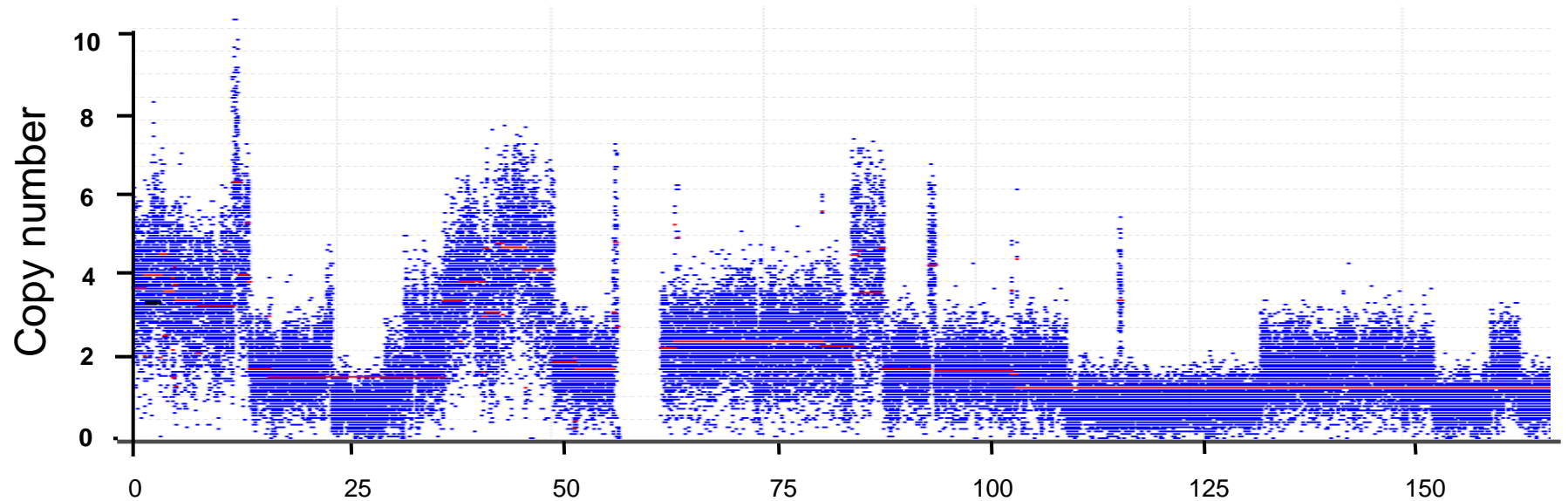
Patterns of somatic structural variation



Patterns of somatic structural variation

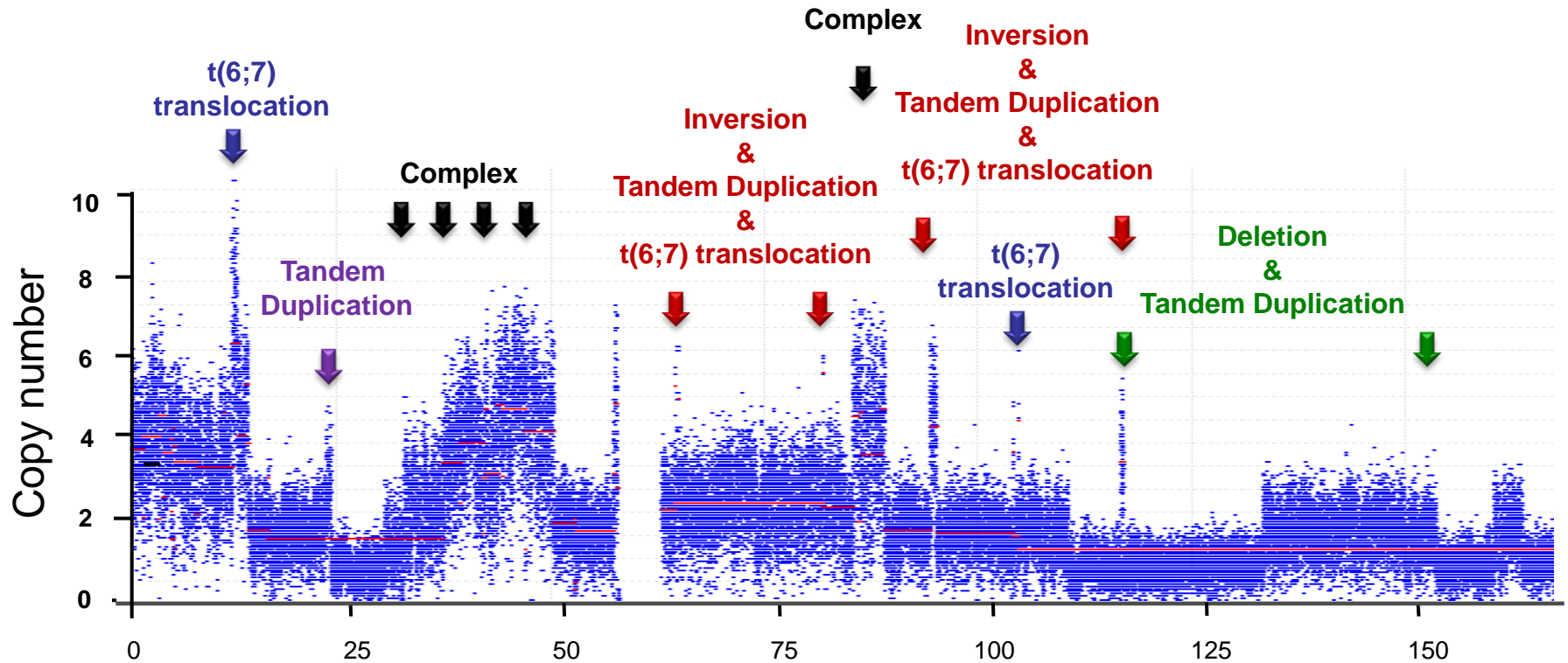


Complex patterns of structural variation



Solexa copy number:- chromosome 6

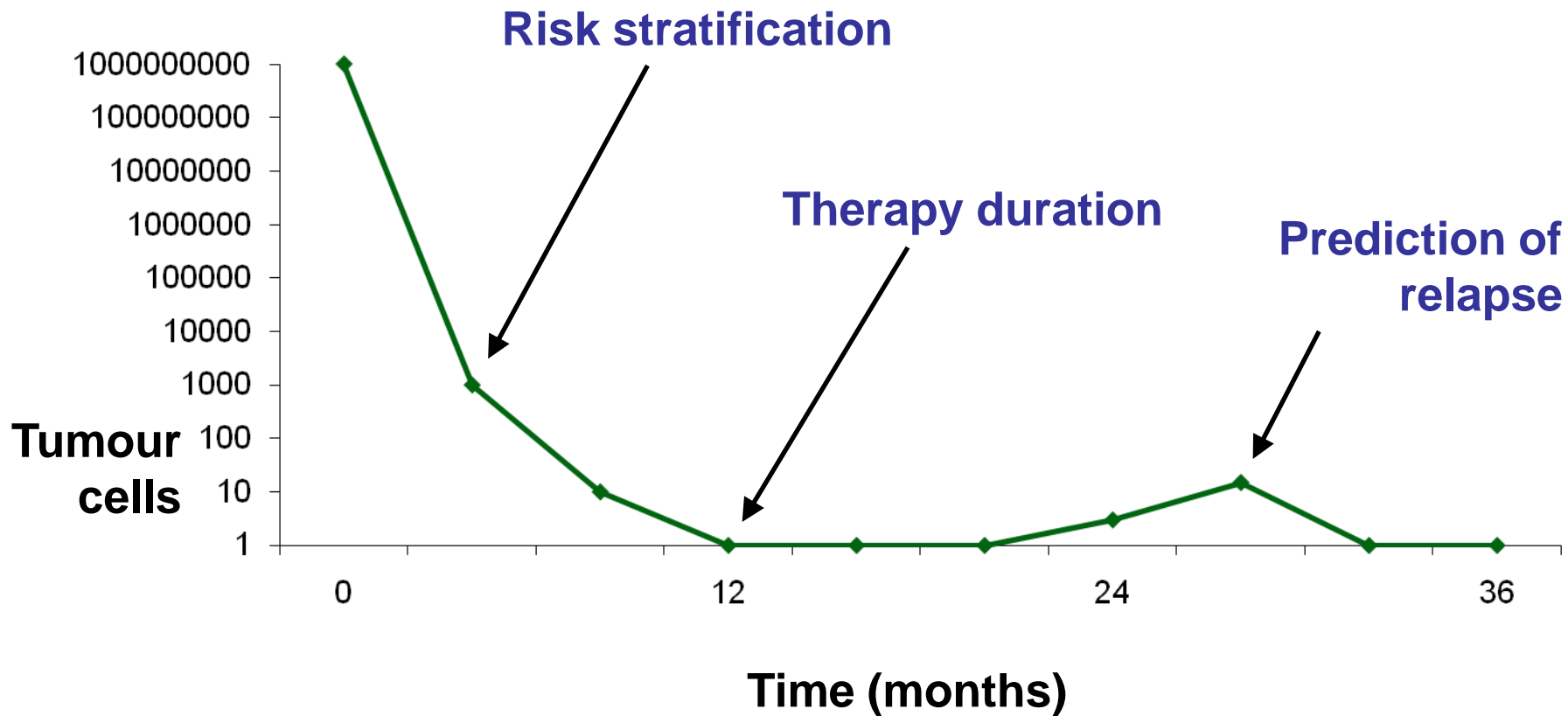
Complex patterns of structural variation



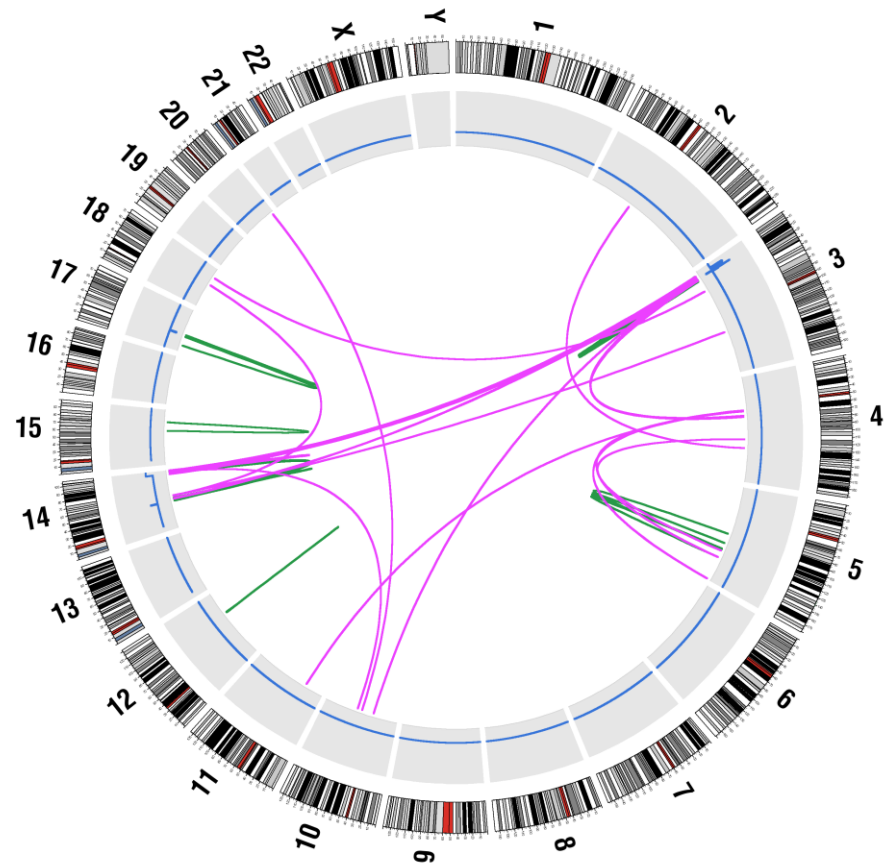
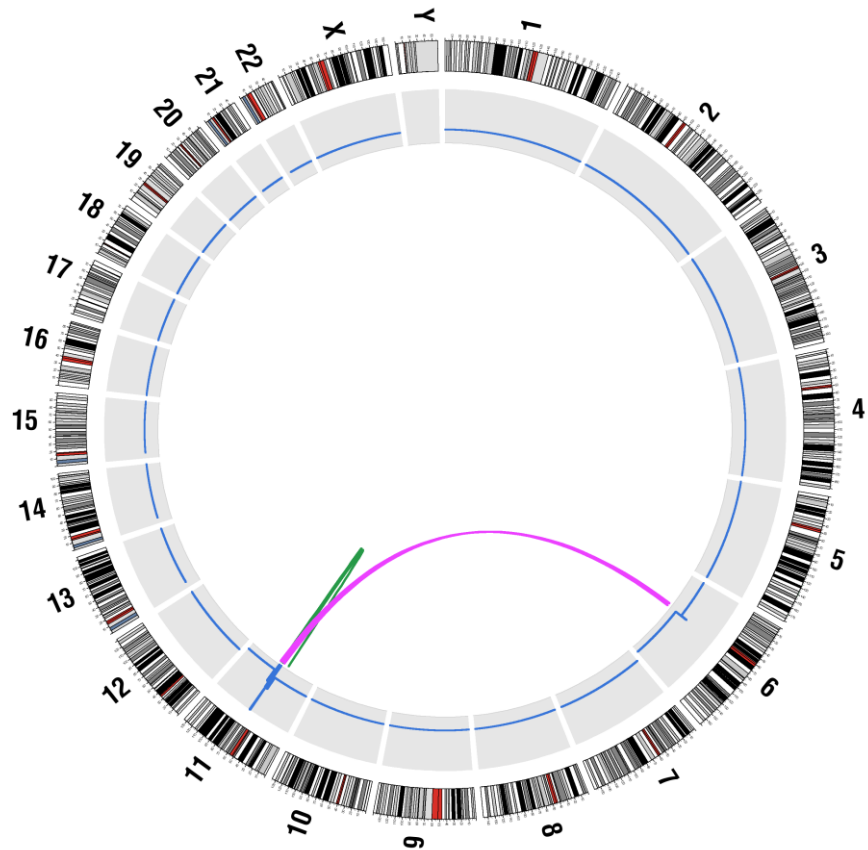
Solexa copy number:- chromosome 6

Potential applications in healthcare

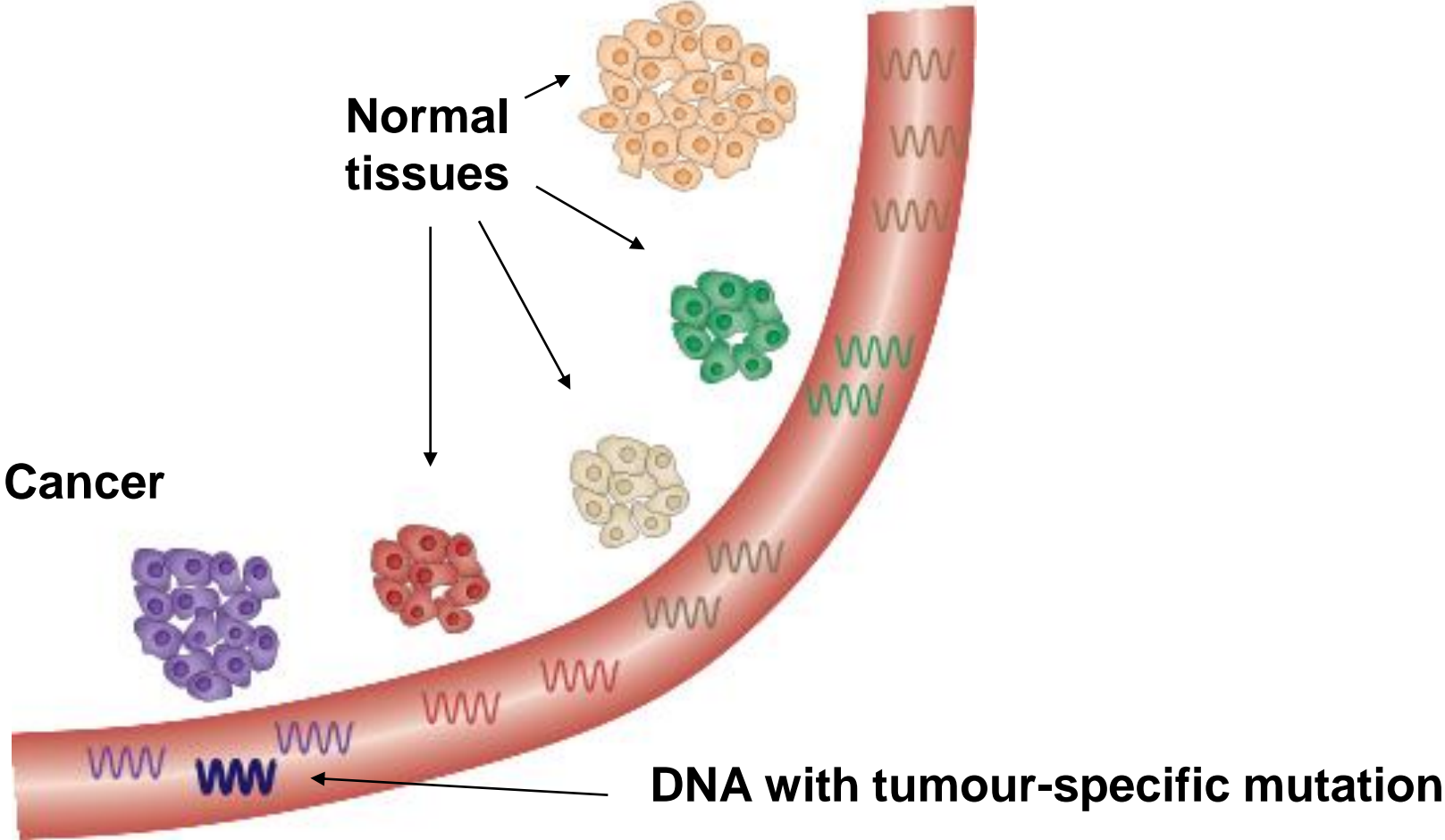
Personalised Haematology



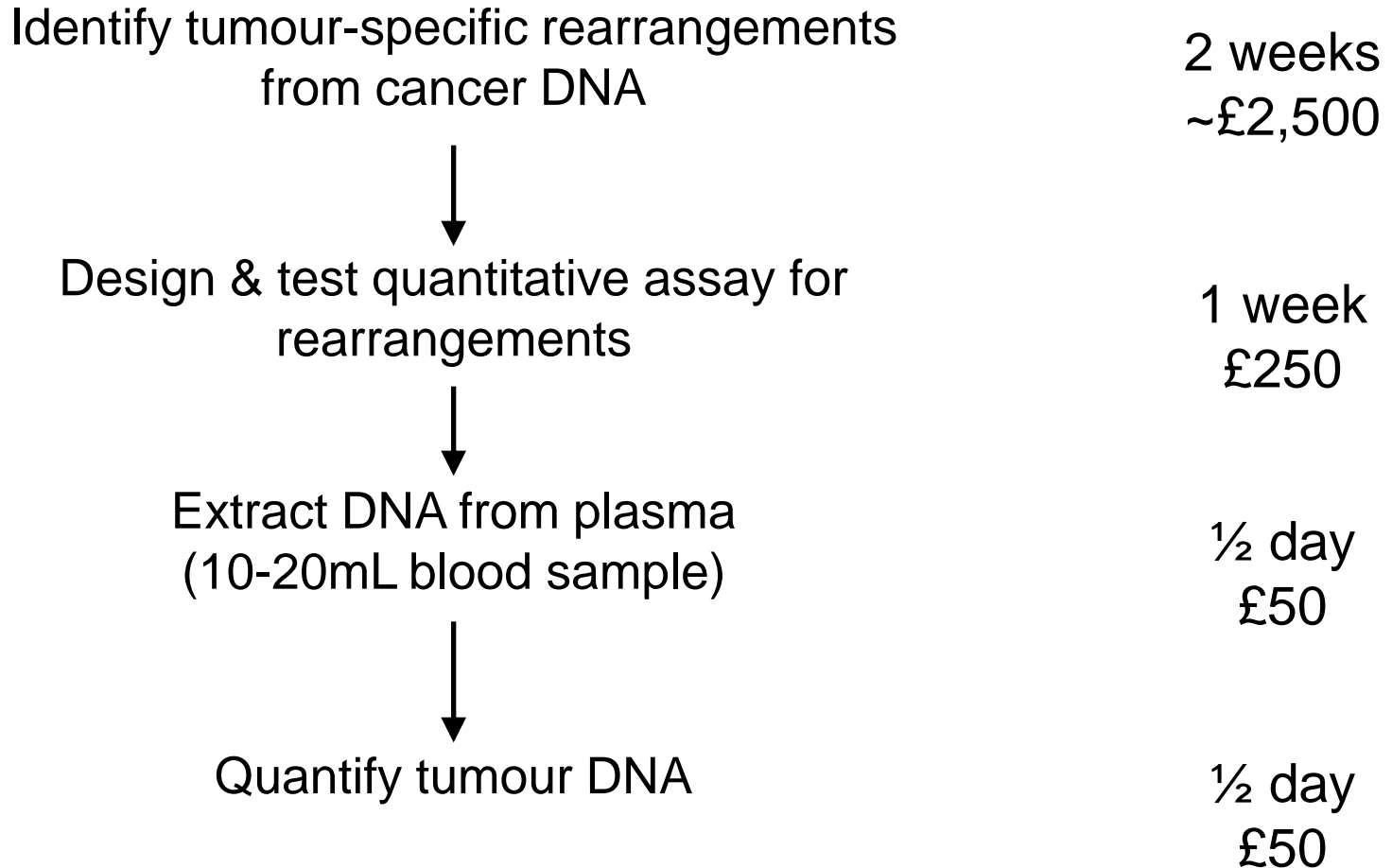
Tumour-specific rearrangements



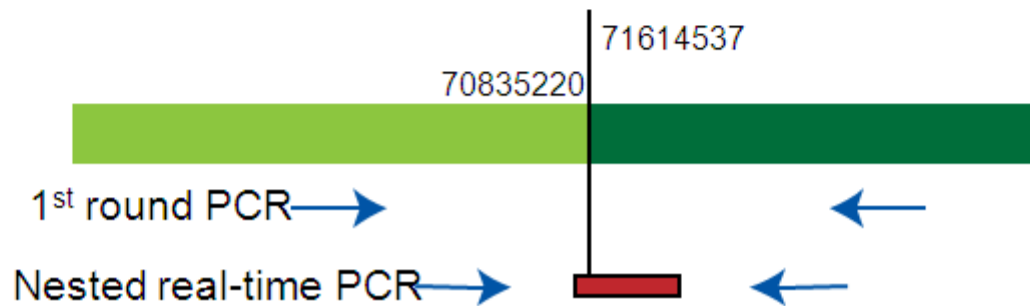
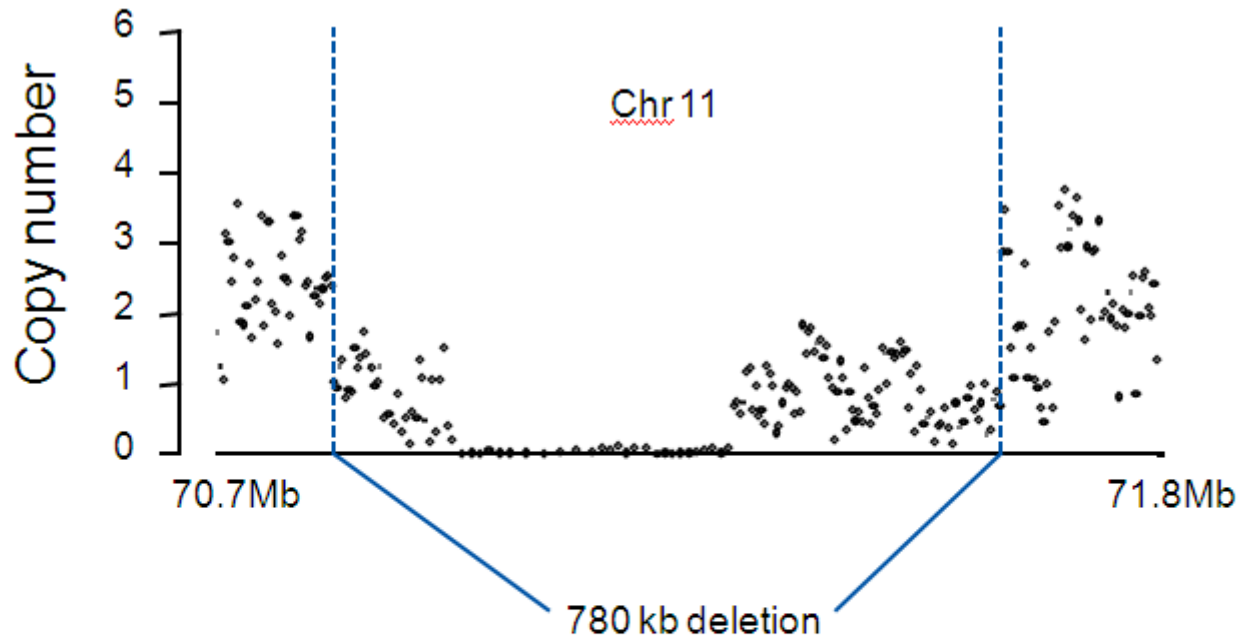
Plasma DNA



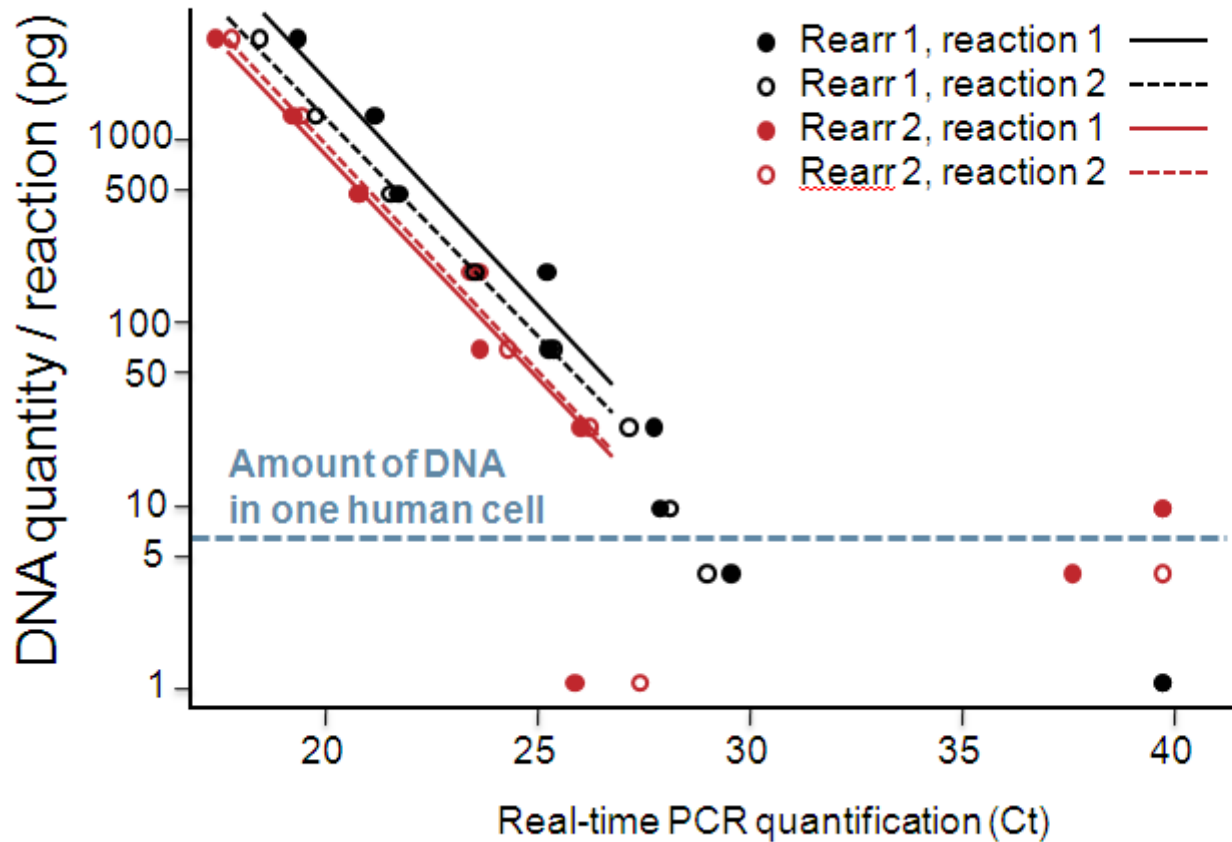
Work flow



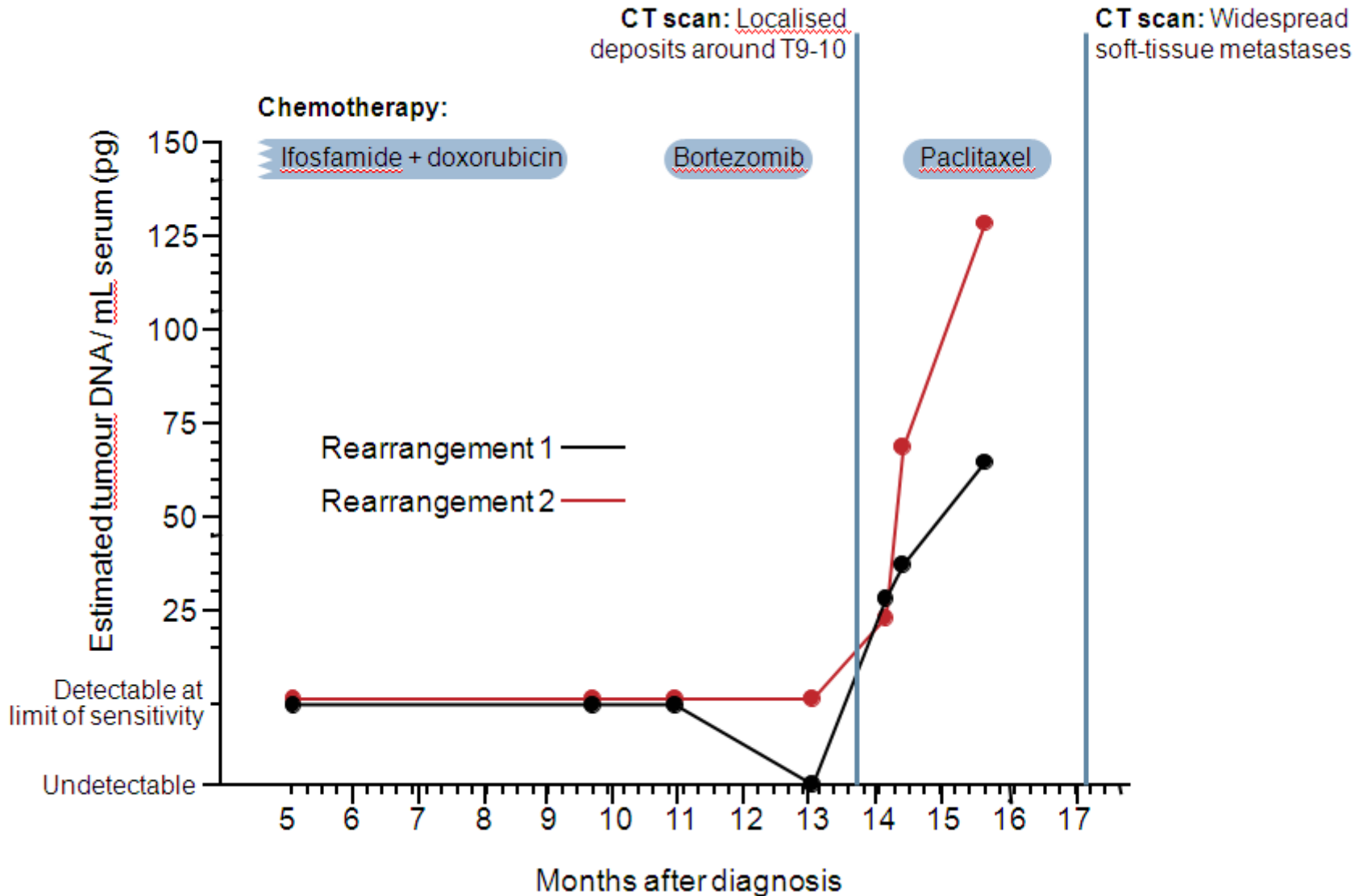
Assay design



Detecting 1 copy of tumour genome



Serial measurements



Potential healthcare applications

- Monitoring tumour response to therapy in real-time
 - Reduce toxicity, prevent drug wastage
- Identifying disease relapse before clinically evident
 - Pre-emptive therapy
- Choosing intensity of adjuvant therapy based on risk stratification
- Surrogate marker of cell kill in early phase clinical trials

Potential healthcare applications

- 100 Breast cancers
- 100 Colorectal cancers
- 100 Osteosarcomas

Sequencing whole cancer genomes

Small Cell Lung Cancer

200,000 cases/year
worldwide

Propensity to be widely
metastatic at diagnosis

2 year survival <15%

Almost exclusively a
disease of smokers

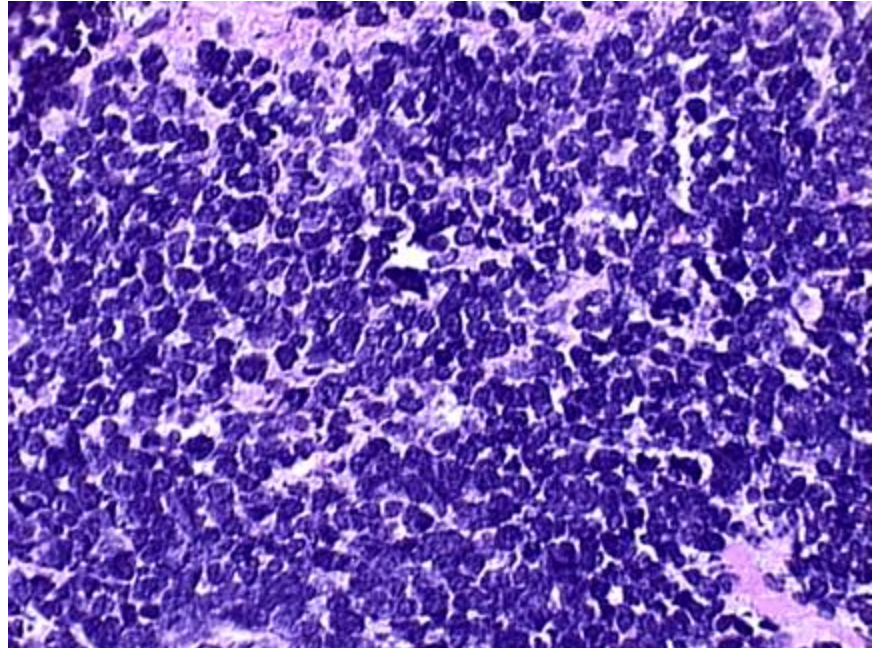
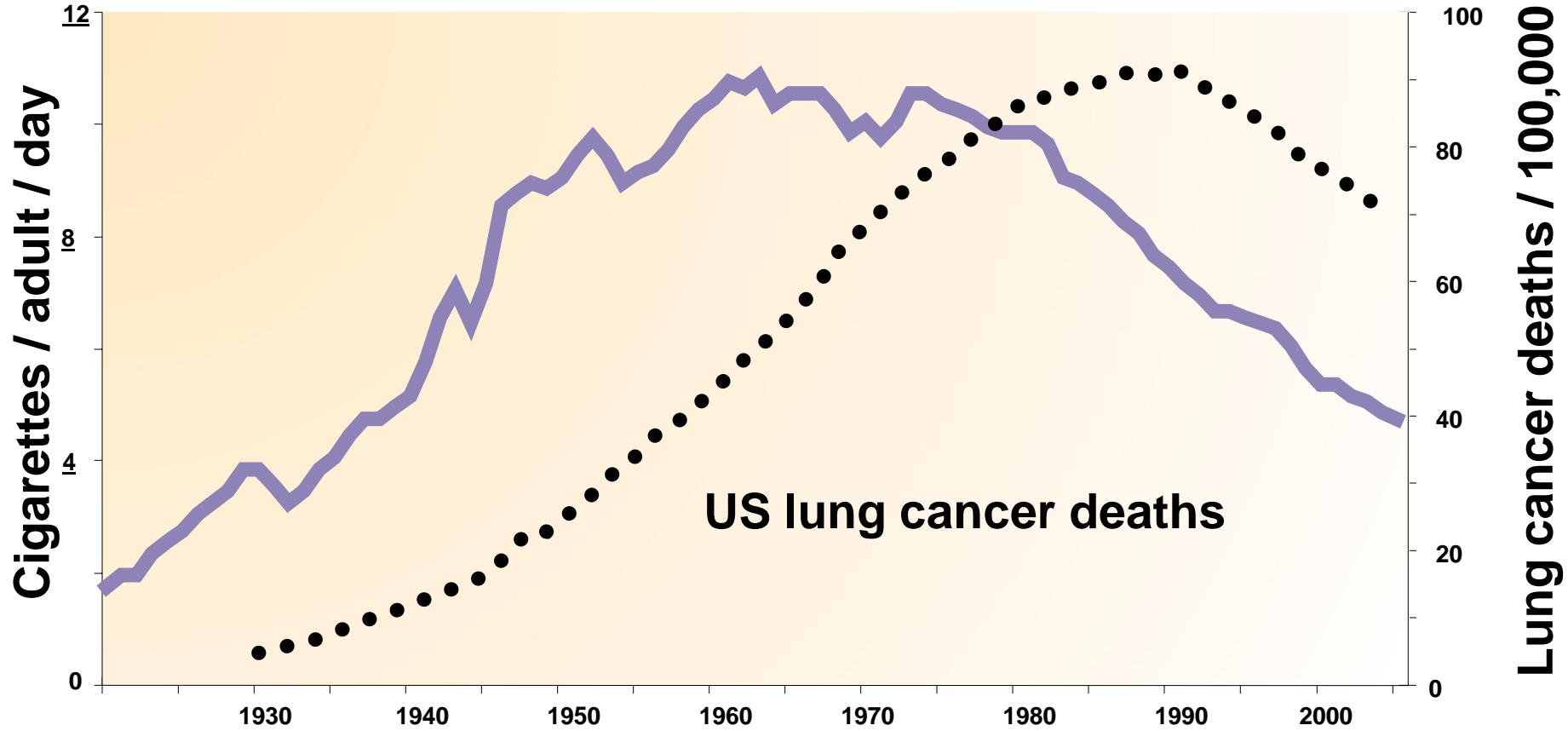
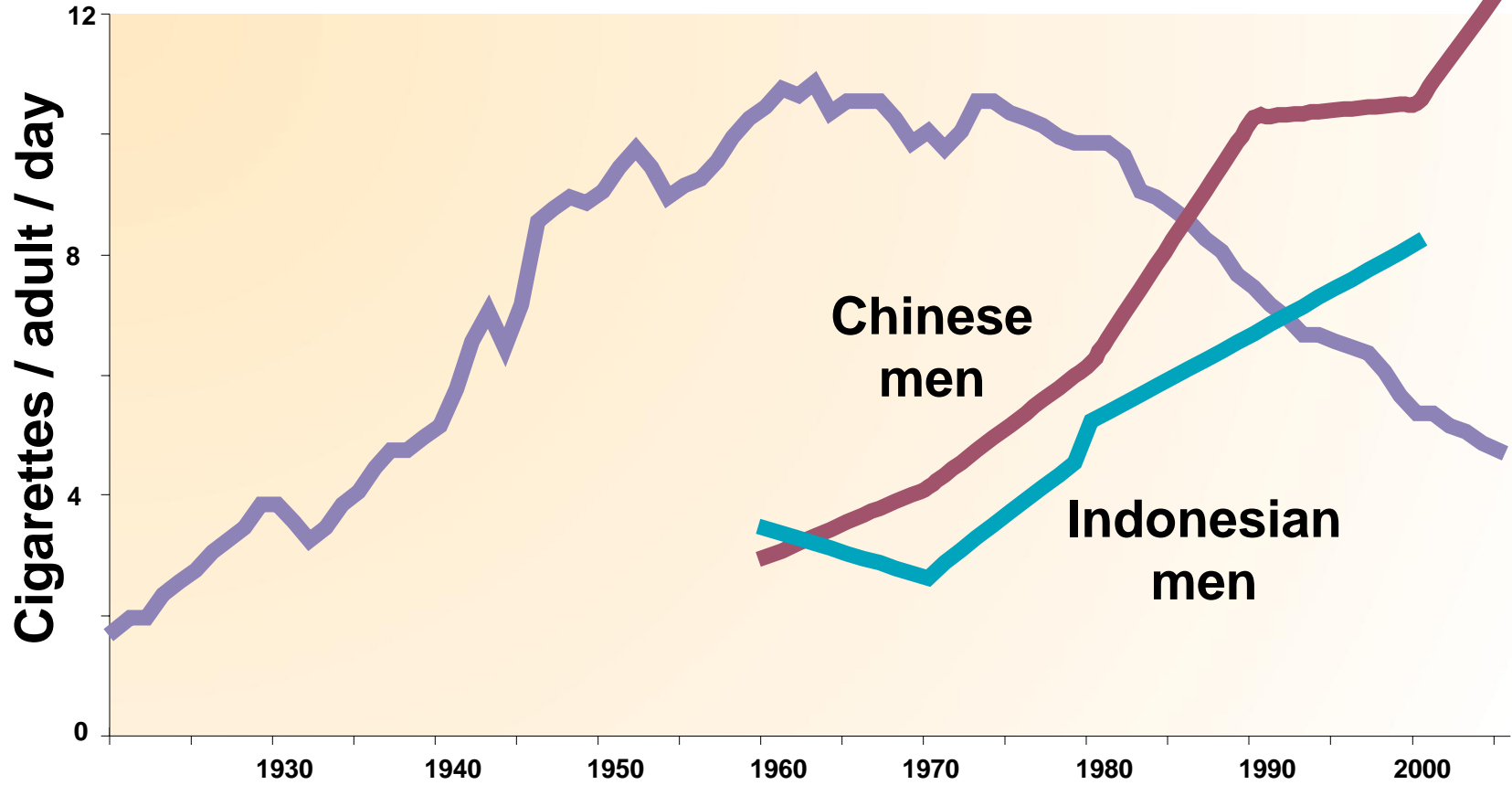


Image from: www.surgical-pathology.com/small_cell_carcinoma.htm

Worldwide smoking trends



Worldwide smoking trends



Cigarette carcinogens

Polycyclic aromatic hydrocarbons



Acrolein

Vinyl chloride

Acetaldehyde

N-Nitrosamines

And >60 others...

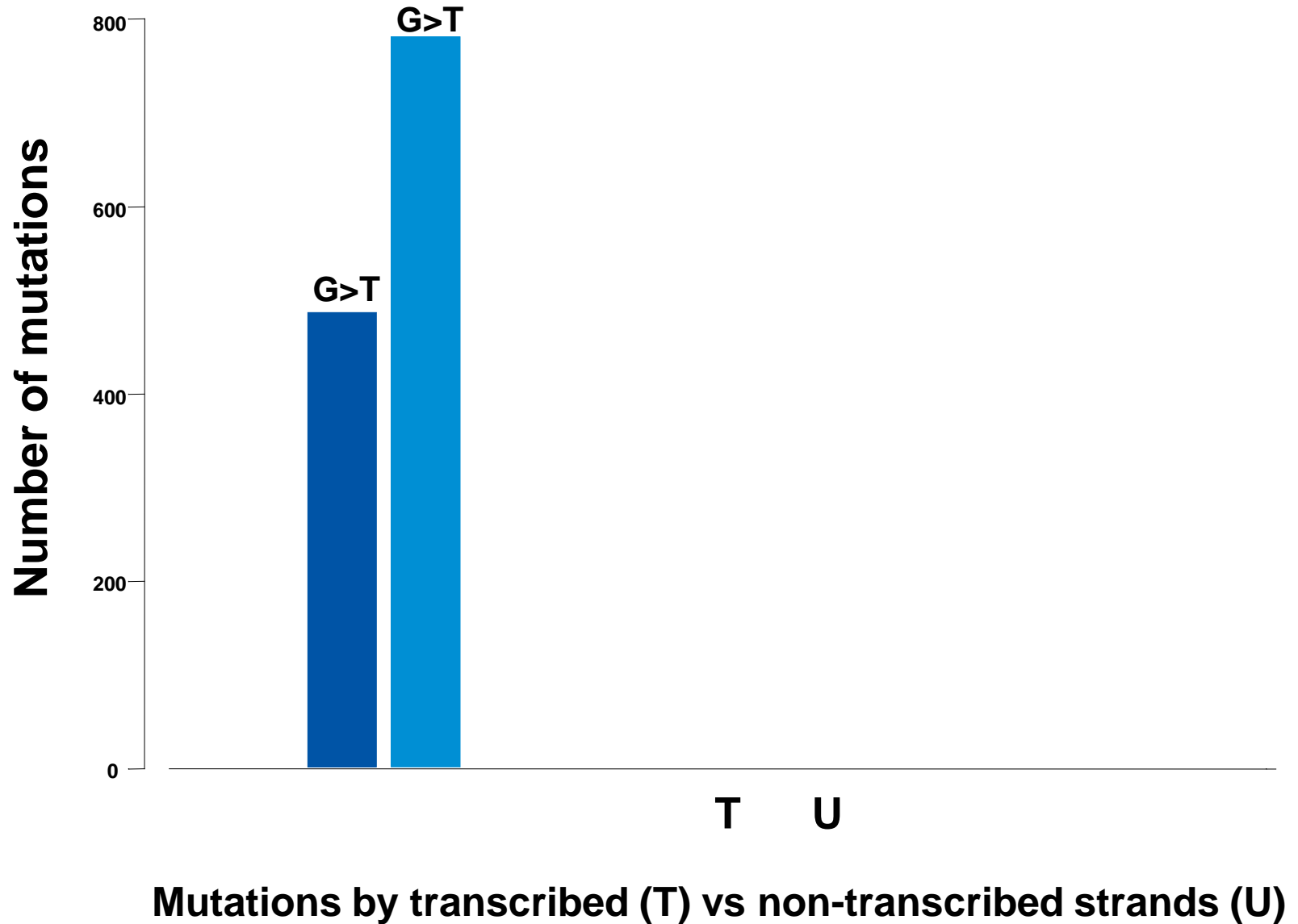
Mutational signatures of tobacco exposure

- Somatic substitutions 22,910
 - Specificity: 97% coding
94% non-coding

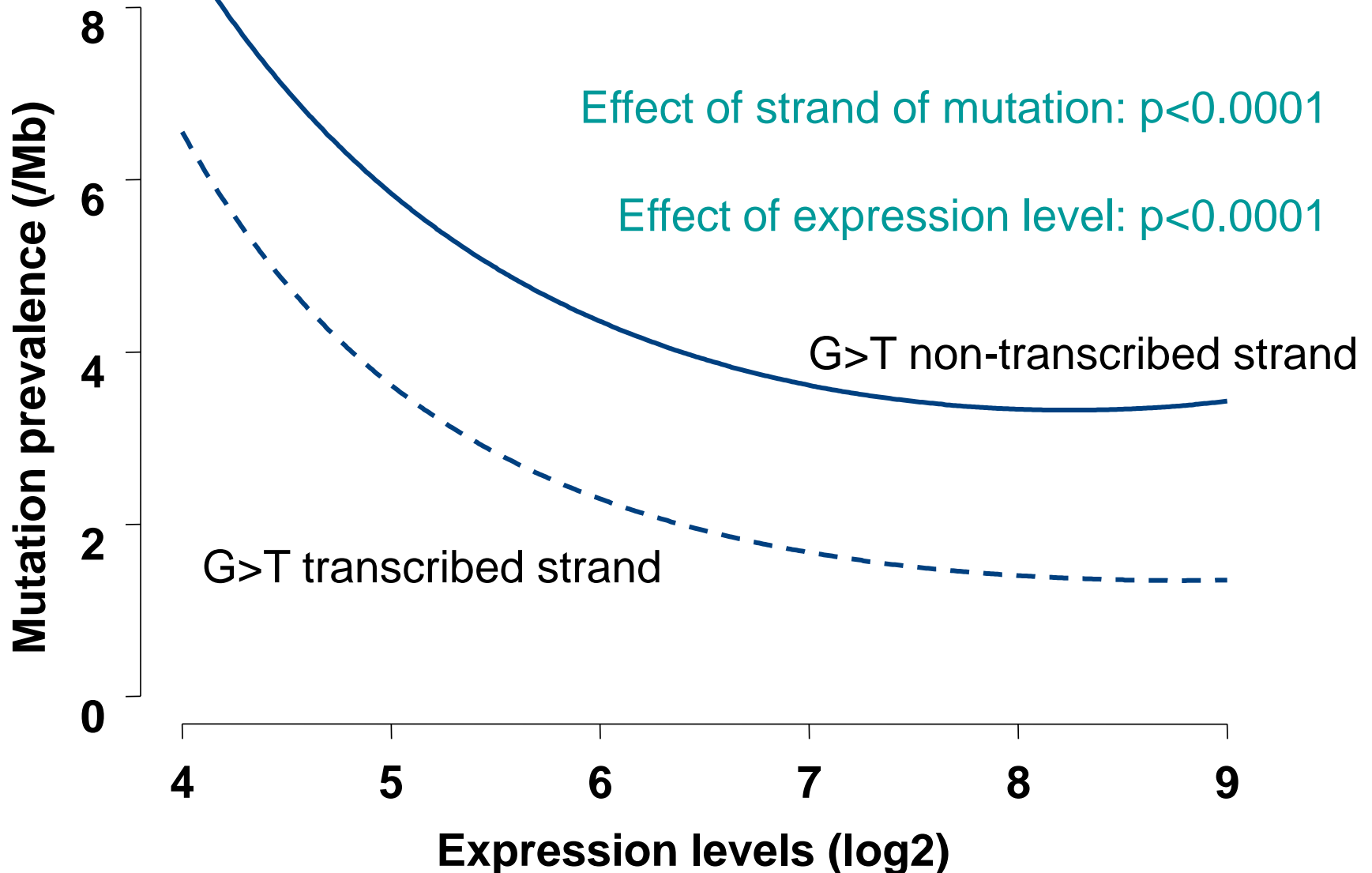
Mutational signatures of tobacco exposure

- Characteristic mutation spectrum
 - A mutations ↓ at GpA (p < 0.0001)
 - G mutations ↑ at CpG (p < 0.0001)
 - G>T & G>A methylated (p < 0.02)
 - G>C unmethylated (p = 0.05)

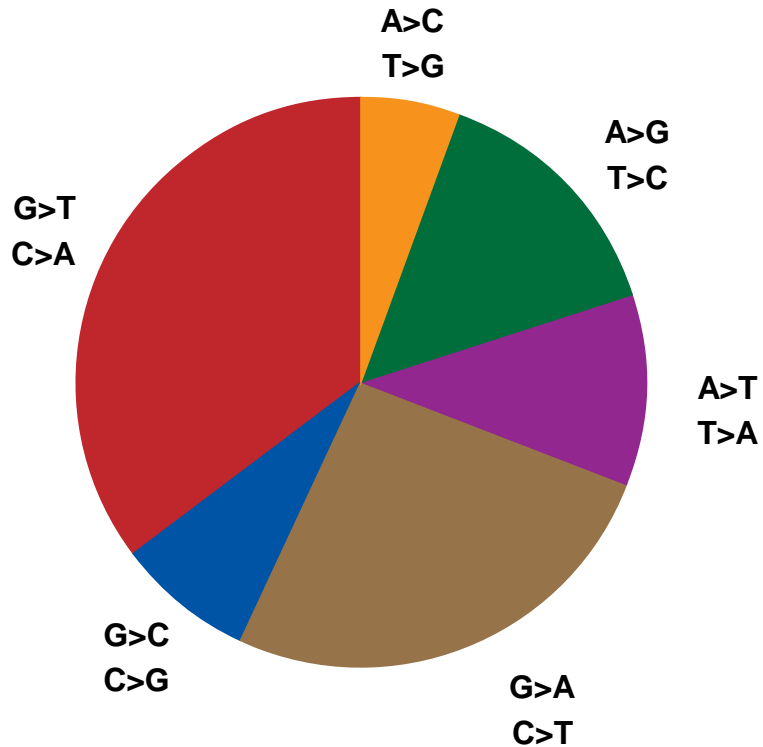
Strand bias in SCLC genome



Transcription-coupled repair and 'expression-linked repair'

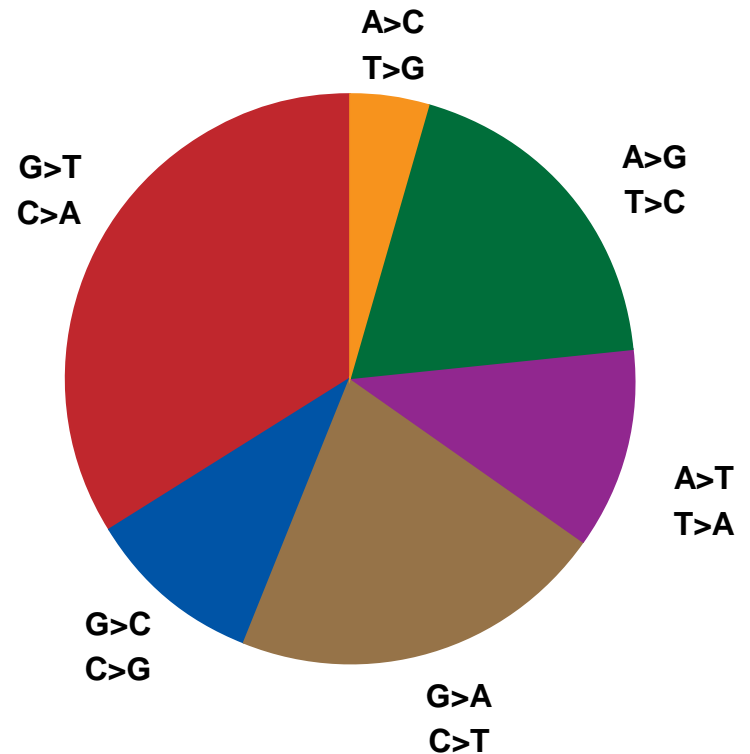


Is this a typical small cell lung cancer?



IARC database: SCLC cases

245 published substitutions in *TP53*



SCLC sequenced

22,910 substitutions genome-wide

Conclusions

Paired-end sequencing is an effective tool for characterising structural variation in complex cancer genomes

The average breast cancer has ~ 100 'somatic' structural variants

The average breast cancer has ~1.0 expressed in-frame fusion gene

Exome sequencing will unveil a multitude of novel drug targets in the next few years

That personalised cancer care is on the horizon



Mike Stratton
Andy Futreal
Peter Campbell

Patrick Tarpey
Stuart McLaren
Adam Butler
Keiran Raine
David Jones
Laura Muddie
Ignacio Varela
Calli Latimer
Serena Nik Zainal

Alison Coffey
Eleanor Howard

Dan Turner
Lira Mamanova

Felix Kokocinski
Carol Scott
Kai Ye



The Wellcome Trust