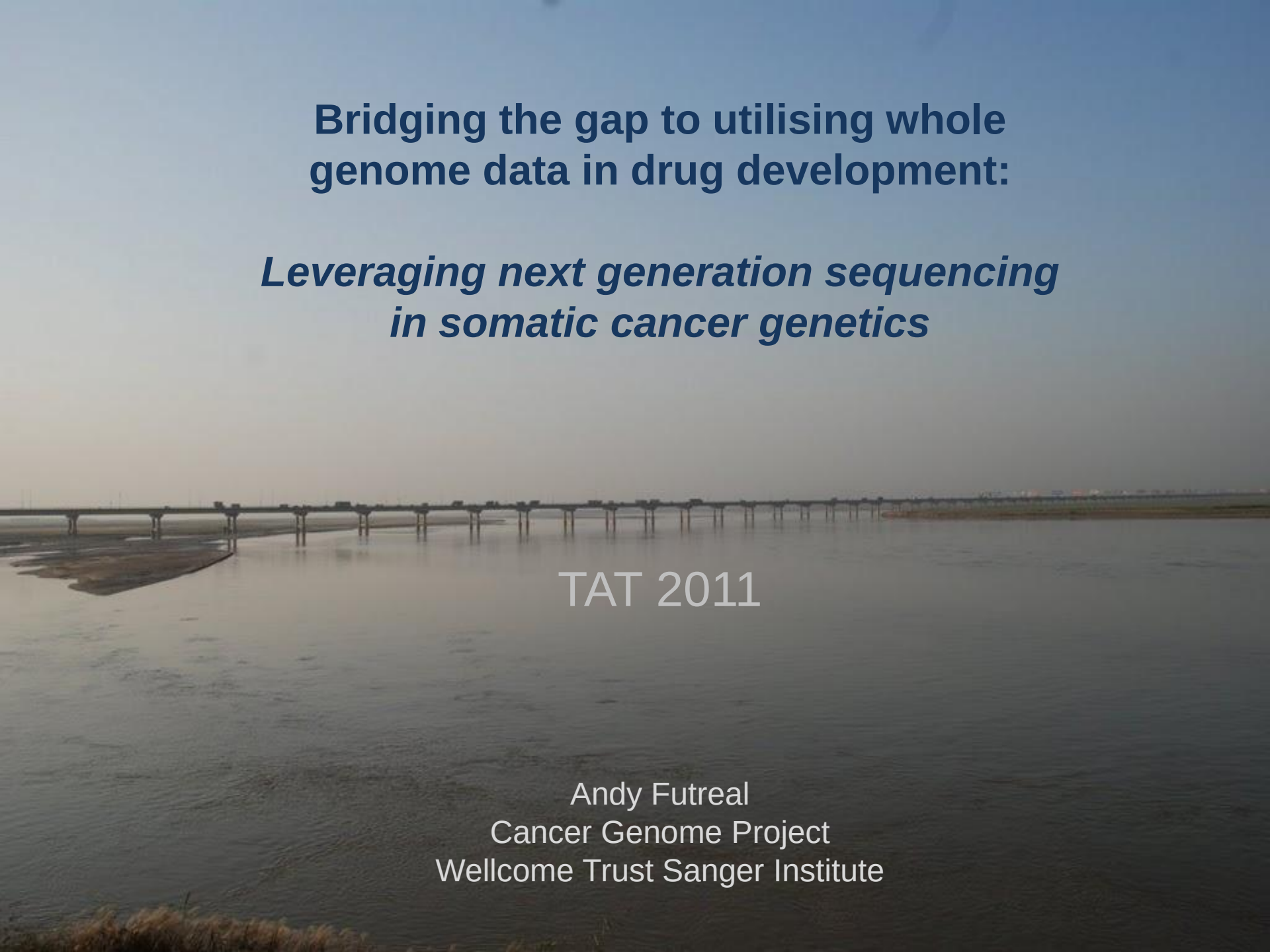


Bridging the gap to utilising whole genome data in drug development:

Leveraging next generation sequencing in somatic cancer genetics

TAT 2011

Andy Futreal
Cancer Genome Project
Wellcome Trust Sanger Institute


wellcome trust


Adenoid Cystic Carcinoma
Research Foundation

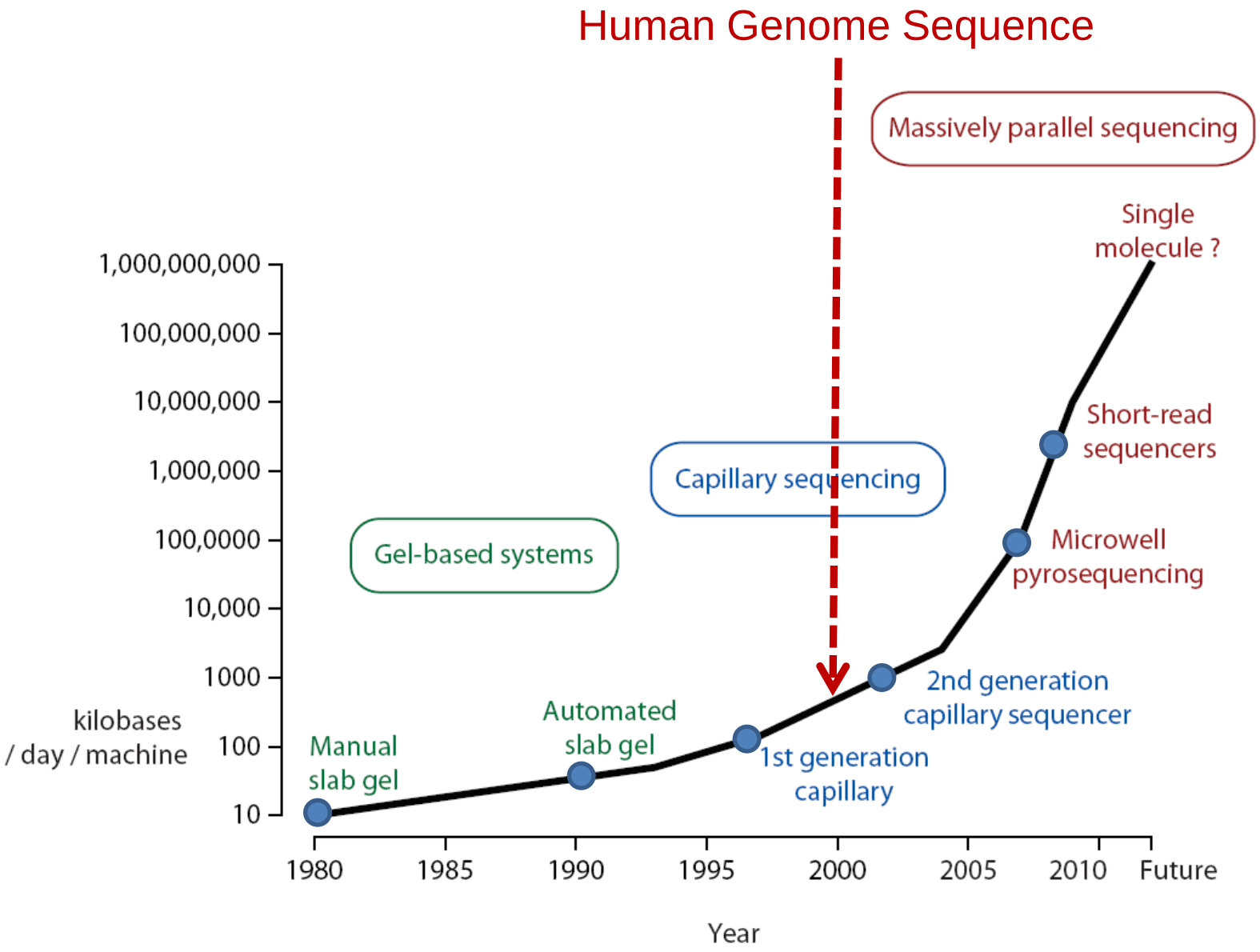

breakthrough
breast cancer

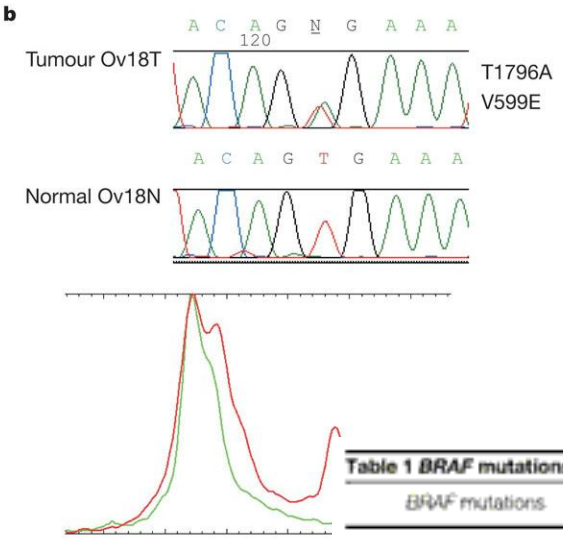

THE KAY KENDALL LEUKAEMIA FUND


The Institute
of Cancer Research


basis
COOPERATION


SEVENTH FRAMEWORK
PROGRAMME
PREDICT
CONSORTIUM


Human Genome Sequence
Massively parallel sequencing
Single molecule ?
Short-read sequencers
Microwell pyrosequencing
2nd generation capillary sequencer
1st generation capillary
Capillary sequencing
Automated slab gel
Manual slab gel
Gel-based systems
1,000,000,000
100,000,000
10,000,000
1,000,000
100,0000
10,000
1000
100
10
kilobases / day / machine
1980
1985
1990
1995
2000
2005
2010
Future
Year



2002 – BRAF mutations in human cancer.

Heteroduplex/PCR resequencing of a handfull of genes

Table 1 *BRAF* mutations in human cancer

BRAF mutations		Cancer cell lines								Primary tumours						
Nucleotide	Amino acid	(1) Mel.	(2) Colo. ca.	(3) Glioma	(4) Lung ca.	(5) Sarcoma	(6) Breast	(7) Ovarian	(8) Other	(1) Mel. STC	(2) Mel.	(3) Colo. ca.	(4) Ovarian*	(5) Sarcoma	(6) Other†	Total
G1388A	G463E							1								1
G1388T	G463V		1													1
G1394C	G465A									1						1
G1394A	G465E										1					1
G1394T	G465V				1											1
G1403C	G468A				2											2
G1403A	G468E											1				1
G1753A	E585K												1			1
T1782G	F594L											1				1
G1783C	G595R		1													1
C1786G	L596V				1											1
T1787G	L596R												1			1
T1796A	V599E	19	5	4		5	1		1	11	5	2	3	1	0	57
TG1796-97AT	V599D	1														1
	Total	20	7	4	4	5	1	1	1	12	6	4	5	1	0	71
No. samples screened		34	40	38	131	59	45	26	172	15	9	33	35	182	104	923
Per cent		59%	18%	11%	3%	9%	2%	4%	0.6%	80%	67%	12%	14%	0.5%	0%	8%

Amino acid residues are grouped in blocks. Three further *BRAF* coding sequence variants were identified (G2041A R681Q in the hEC1A endometrial cancer cell line, T974C I325T in the ZR-75-30 breast cancer cell line, and C2180T A727V in the Ht33AJ-JA1 T-ALL cell line). These were not present in 341 control DNAs. Lane numbers (in parentheses) are provided for convenience. Mel., melanoma; Colo. ca., colorectal cancer; Mel. STC, melanoma short-term culture.

*Four out of ten LMP (low malignant potential); 1 out of 25 malignant epithelial.

†Glioma ($n = 15$), breast cancer ($n = 33$), prostate cancer ($n = 23$), HNSCC (head and neck squamous cell carcinoma) ($n = 19$), lung cancer ($n = 14$).

What we learned from earlier PCR- based systematic screens

- The majority of somatic mutations identified in large-scale resequencing screens are likely to be passenger events
- There is evidence for multiple infrequently mutated genes under positive selection
- There are likely to be many genes that can contribute to oncogenesis when mutated
- A large number of genes sequenced in a large number of cancers of “same” type will be needed to begin to fully elucidate the complement of cancer genes

PCR exon-resequencing ccRCC study

101 ccRCC cases

(96 clinical samples + 5 matched pair cell lines

Screen the coding exons of 3,544 genes

(750 Mb total)

Copy number and expression

Follow-up series of 311 clear cell carcinoma clinical samples with matching normals

Results

VHL point mutations in ~60%, VHL/Hypoxia signature in ~85%

Very quiet on SNP6.0, no high level focal amplicons

75/91 (82%) of cases assessed for expression had upregulation of genes associated with cellular hypoxia

Four/five significantly mutated genes in ccRCC are histone methylase/demethylases

UTX – H3K27 demethylase and MLL2/3 complex in H3K4 methylation

12/407 cases mutated (12/12 truncating)

SETD2 – H3K36 methyltransferase

15/407 cases mutated (12/15 truncating)

JARID1C – H3K4 demethylase

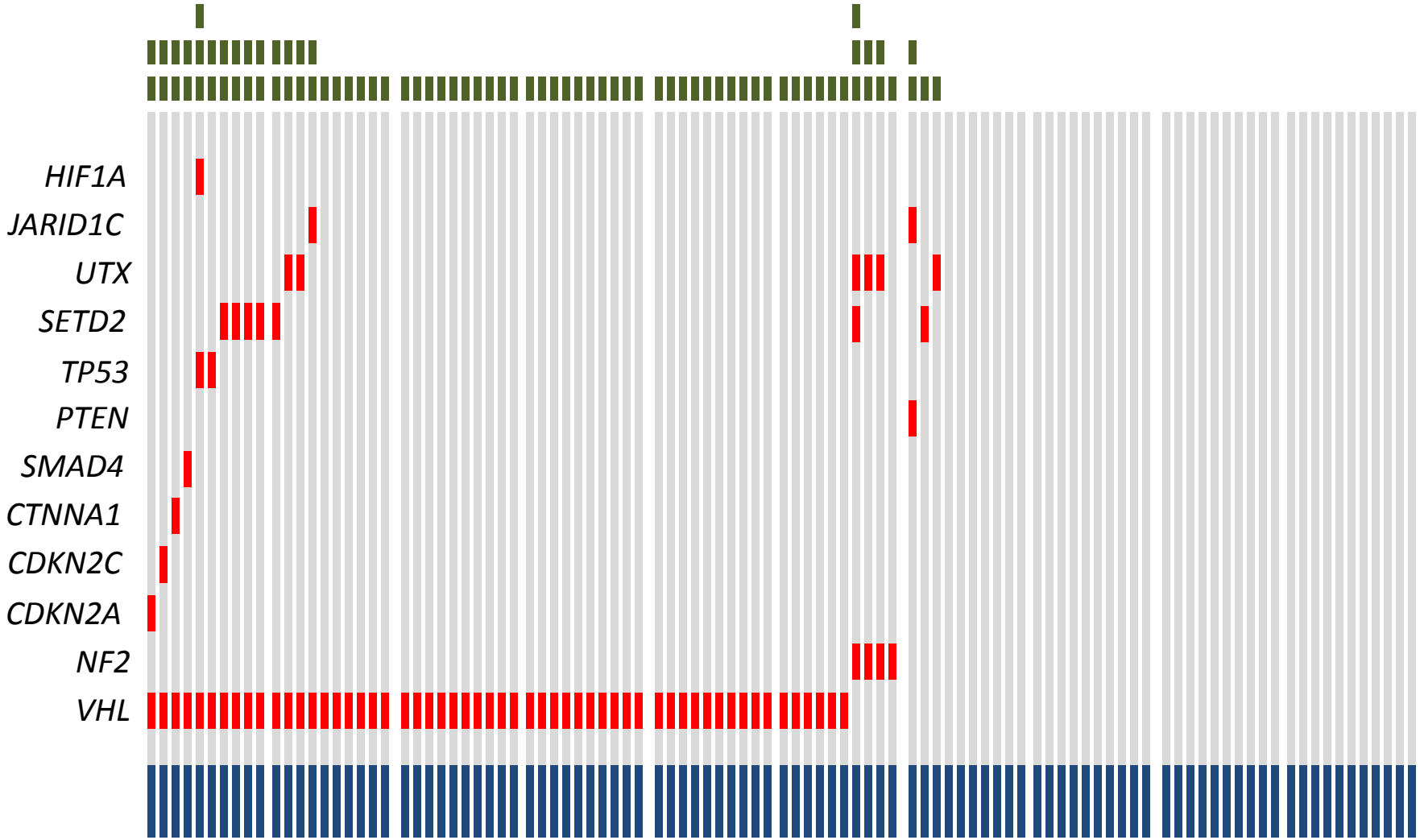
14/407 cases mutated (12/14 truncating)

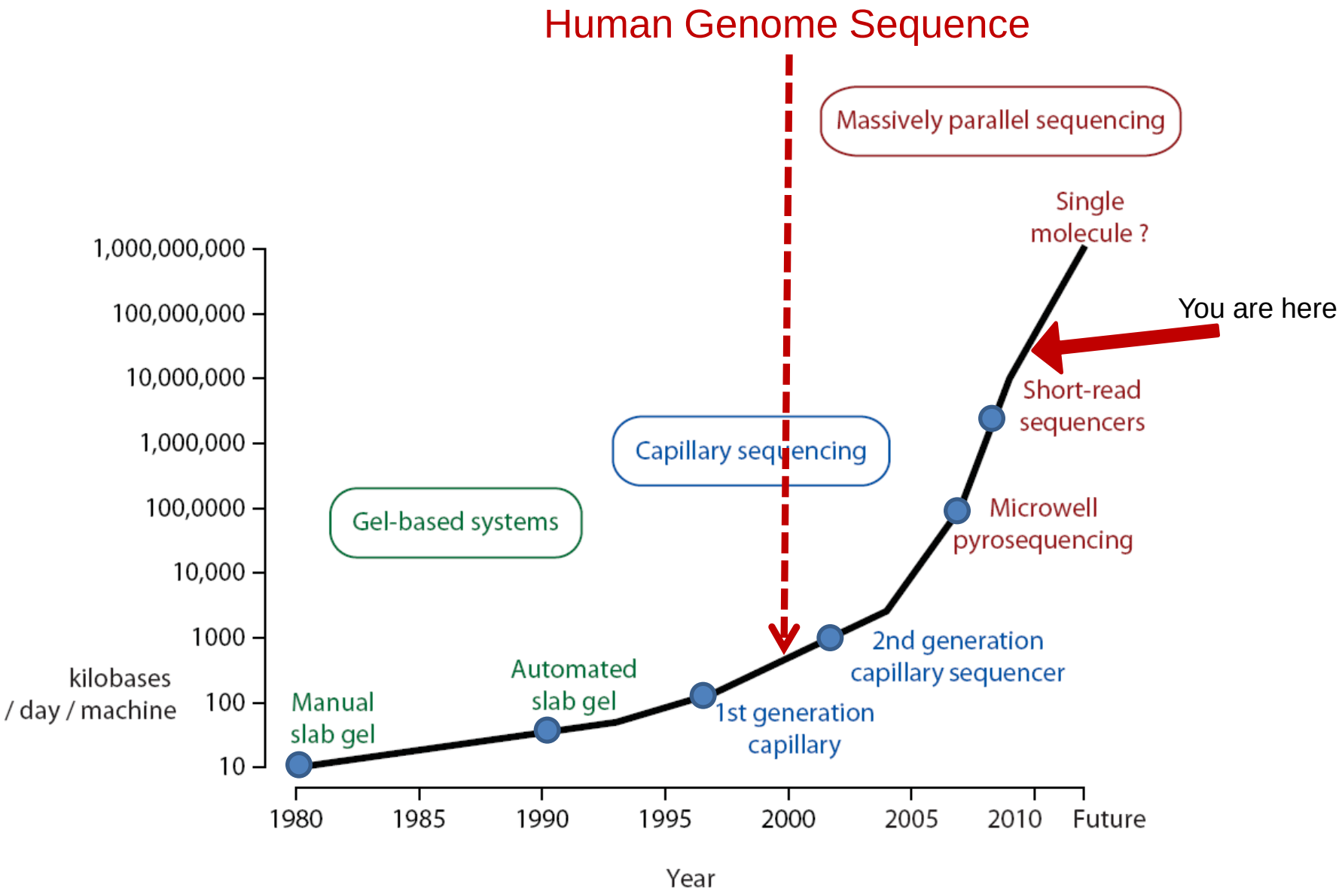
MLL2 – H3K4 methyltransferase

17/407 (including a silent, 6 truncating, 5 missense)

van Haaften et al. Nat Genet. 2009 May;41(5):521-3. Epub 2009 Mar 29.
Dalgliesh et al. Nature. 2010 Jan 21;463(7279):360-3. Epub 2010 Jan 6.

The allelic architecture of driver mutations in clear cell renal cancer

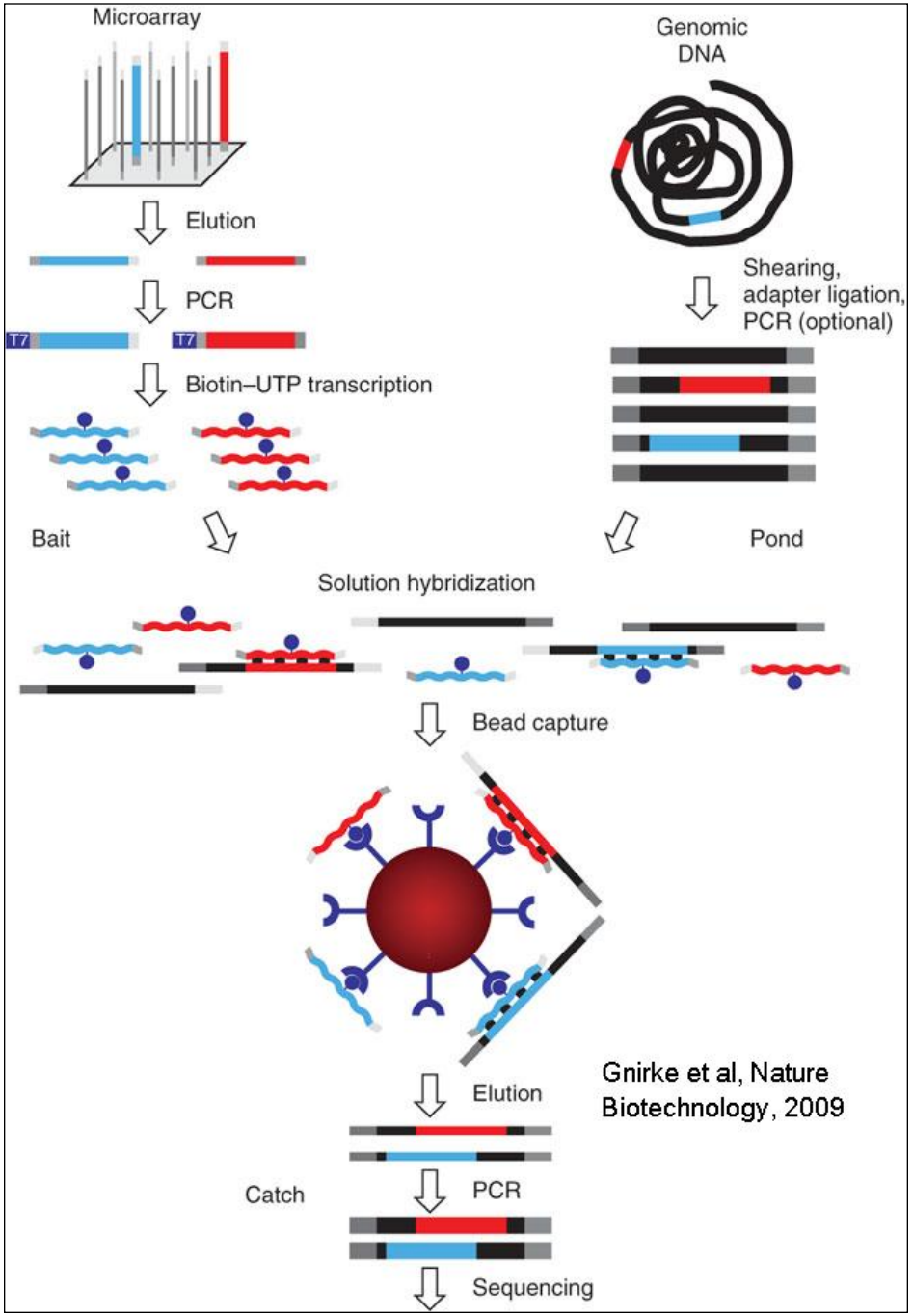



Human Genome Sequence
Massively parallel sequencing
Single molecule ?
You are here
Short-read sequencers
Microwell pyrosequencing
2nd generation capillary sequencer
1st generation capillary
Capillary sequencing
Automated slab gel
Manual slab gel
Gel-based systems
1,000,000,000
100,000,000
10,000,000
1,000,000
100,0000
10,000
1000
100
10
kilobases / day / machine
1980
1985
1990
1995
2000
2005
2010
Future
Year

Whole ‘exome’ sequencing

Rearrangement screens

Whole genome shotgun



Solution Hybrid Capture of all coding exons and miRNA genes

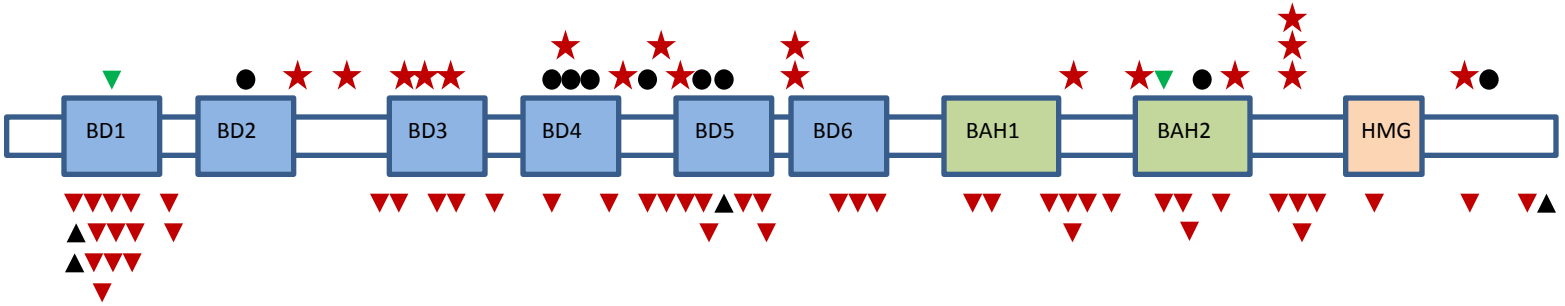
Exome is GenCode/ICGC includes 21,416 protein coding genes + 1664 miRNA genes

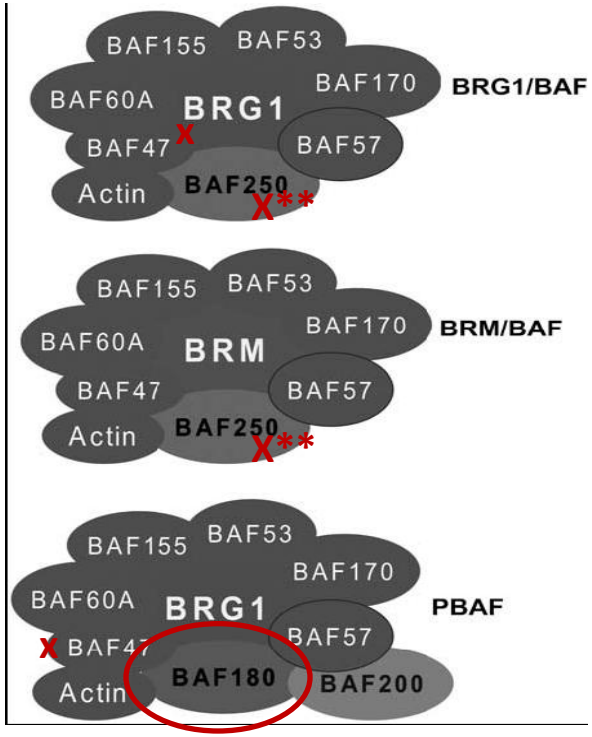
Clinical Samples in exome sequencing

Sample	Sex	Age	Grade	Histology	VHL mutation^	SETD2 mutation	UTX mutation
PD2125a	M	82	4	Clear Cell			
PD2126a	F	74	1	Clear Cell	c.236_241delGCAGTC; p.R79_P81>P	c.1801T>A; p.R601*	
PD2127a	F	59	4	Clear Cell			
PD2144a	F	63	4	Clear Cell	c.525delC; p.Y175*		
PD2147a	F	50	2	Clear Cell			c.4161_4162delTG; p.Y1387fs*1
PD3295a	M	62	4	Clear Cell			
PD3441a	M	69	1	Clear Cell	c.223_225delATC; p.F76_C77>C		

^ VHL mutations in PD2126a and PD3441a were not "re-discovered" in exome sequencing due to poor coverage of the highly GC-rich first exon.

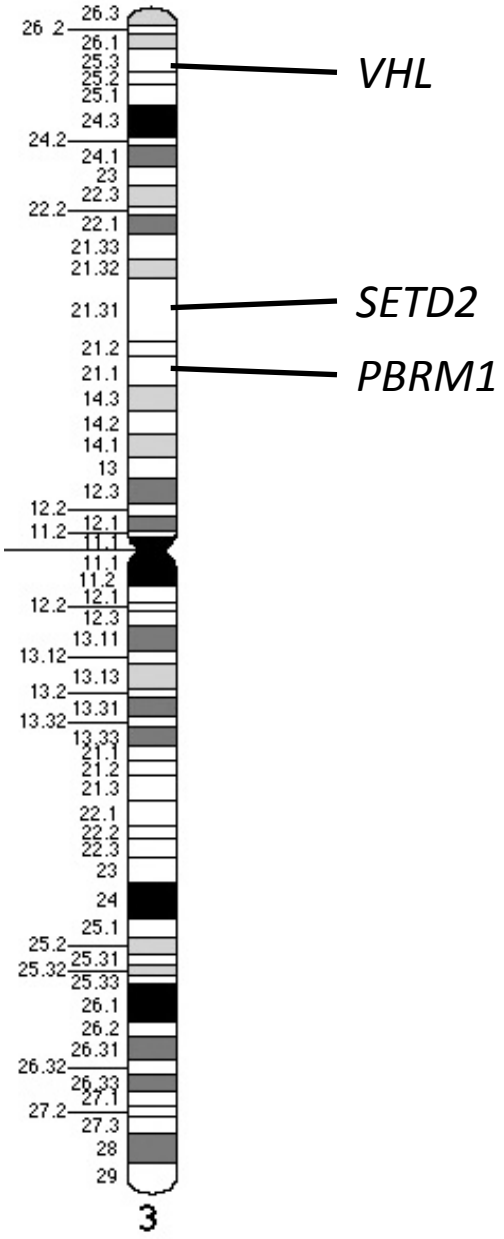
PBRM1 is somatically mutated in 40% (92/227) of ccRCC





ARID1A ** also mutated in ccRCC

Nature, Jan 20, 2011



36/38 PBRM1 mutant ccRCC have hypoxia signature

55/107 cases with a demonstrable* VHL mutation have a PBRM1 mutation

9/9 cases with a SETD2 mutation have mutation in either VHL or PBRM1

6/9 SETD2 mutant cases have a mutation in all three gene

3 tumour suppressor genes unmasked with only 4 hits

3p LOH is most frequent marker in ccRCC

*point mutations only

Breast Cancer Exome Sequencing

- 72 cases
  - 46 ER+ (HER2-)
  - 11 Triple Negative
  - 12 HER2+
  - 1 BRCA1 mut, 2 BRCA2 mut

- Exome is GenCode/ICGC includes 21,416 protein coding genes + 1664 miRNA genes

Solution hybrid capture (Agilent) followed by sequencing on Illumina and variant calling as just described

Follow-up in 300 cases

ER+ Cases I

Sample	TOTAL mutations	AKT1	AKT2	CDH1	GATA3	KRAS	MAP2K4	NF1	PIK3CA	PTEN	RB1	SETD2	STK11	SMAD4	TP53
PD3852a	13														
PD3854a	20														
PD3856a	50								PIK3CA						TP53
PD3857a	22								PIK3CA						
PD3858a	61														
PD3985a	33								PIK3CA						TP53
PD3986a	24														TP53
PD3989a	25								PIK3CA	PTEN					
PD3993a	24														
PD3994a	7														
PD3995a	39	AKT1			GATA3		MAP2K4	NF1							
PD3996a	14									PTEN					
PD3997a	13														
PD3999a	28														
PD4000a	48														
PD4004a	44				GATA3										
PD4087a	44				GATA3										
PD4092a	25														
PD4093a	78														TP53
PD4094a	86								PIK3CA						
PD4101a	19														
PD4104a	36				GATA3					PTEN					
PD4105a	40				GATA3										TP53
PD4108a	26								PIK3CA						
PD4110a	25			CDH1					PIK3CA						
PD4111a	18				GATA3										
PD4134a	51		AKT2												

Triple Negative Cases

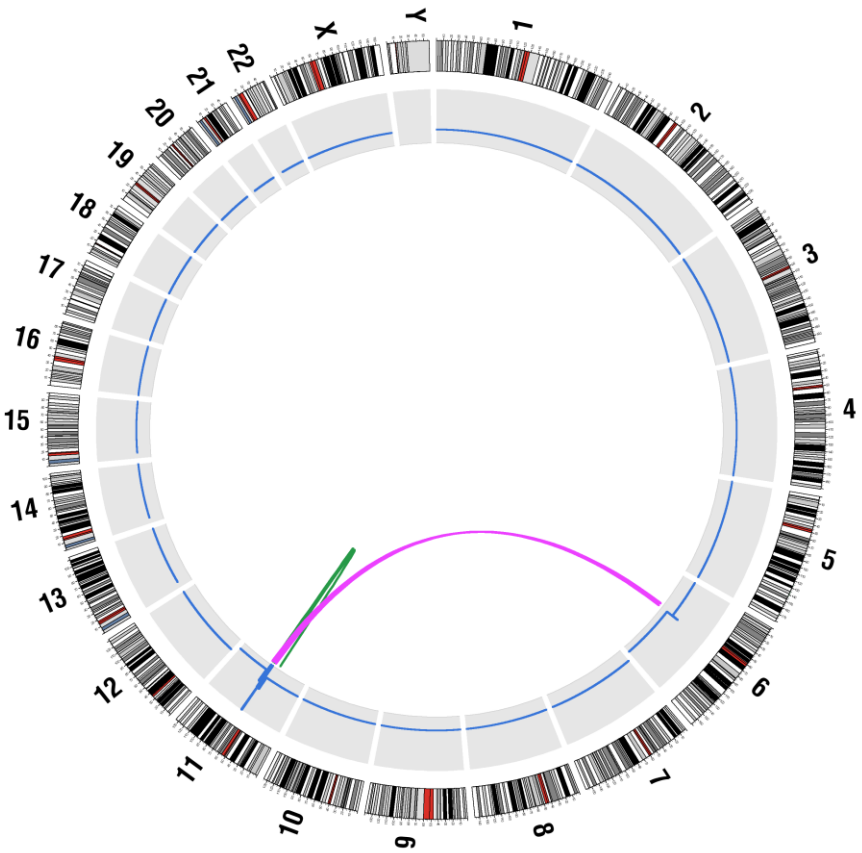
Sample	TOTAL mutations	AKT1	AKT2	CDH1	GATA3	KRAS	MAP2K4	NF1	PIK3CA	PTEN	RB1	SETD2	STK11	SMAD4	TP53
PD3987a	**38**														**TP53**
PD4002a	**118**														**TP53**
PD4003a	**126**														
PD4091a	**36**														**TP53**
PD4098a	**106**														**TP53**
PD4102a	**77**														**TP53**
PD4107a	**83**														**TP53**
PD4109a	**121**														**TP53**
PD4113a	**39**														**TP53**
PD4130a	**35**														**TP53**
PD4133a	**88**										**RB1**				**TP53**

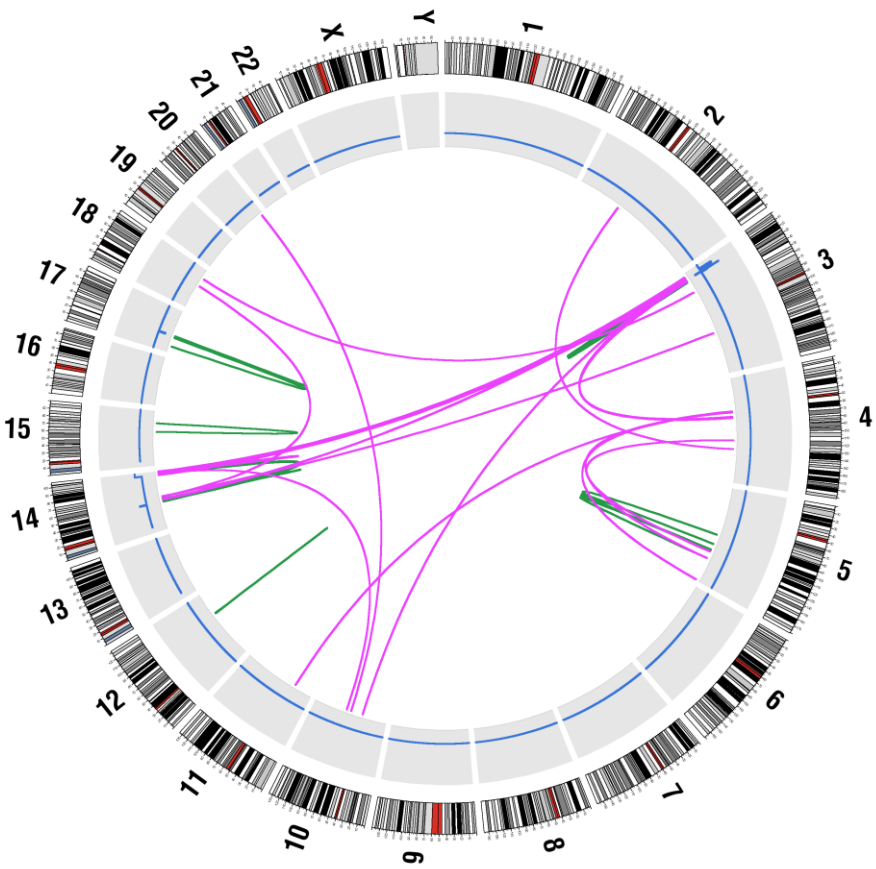
Whole ‘exome’ sequencing

Rearrangement screens

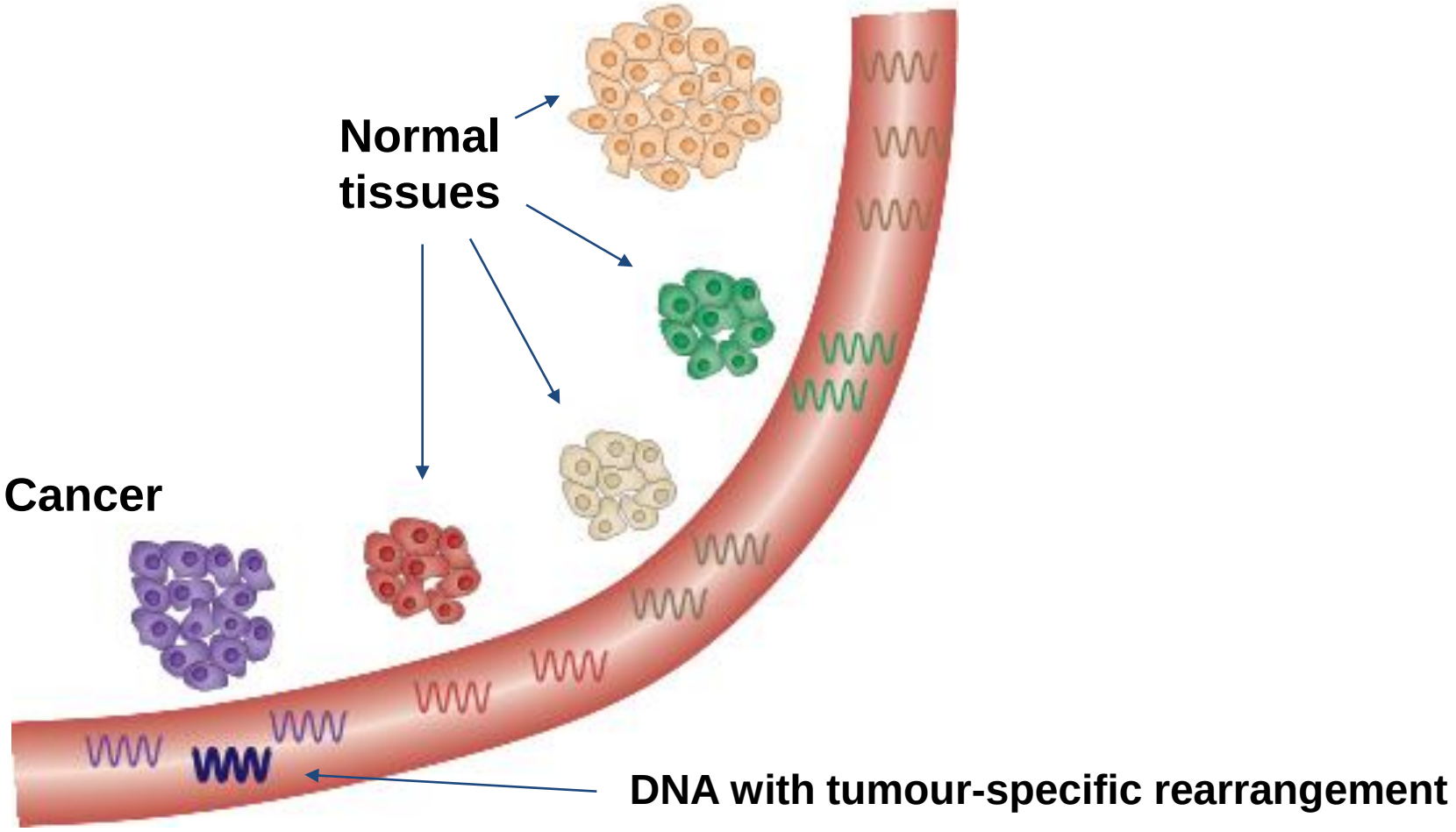
Whole genome shotgun

Tumour-specific rearrangements

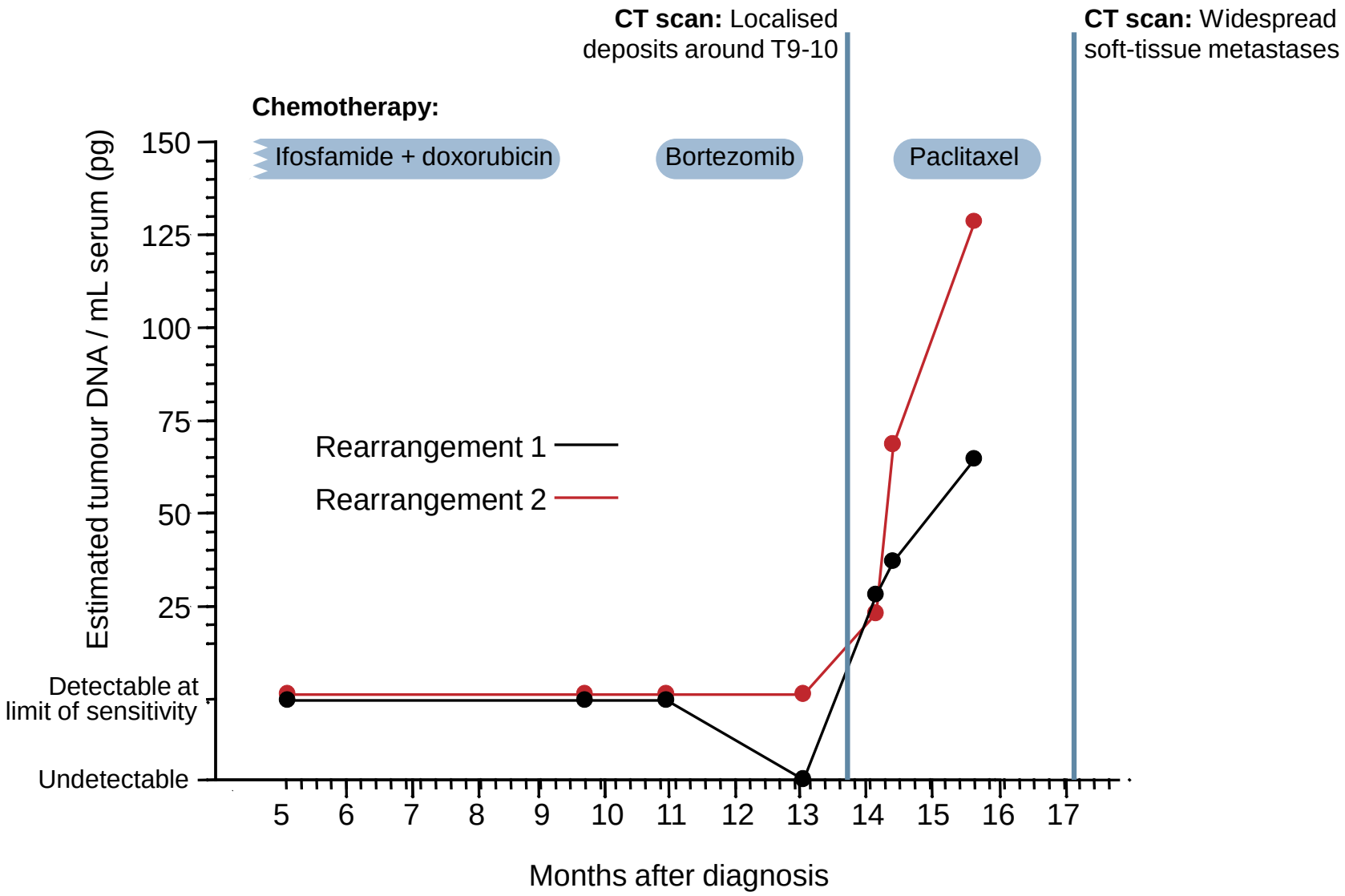




Plasma DNA


Normal tissues
Cancer
DNA with tumour-specific rearrangement

Serial measurements



Whole ‘exome’ sequencing

Rearrangement screens

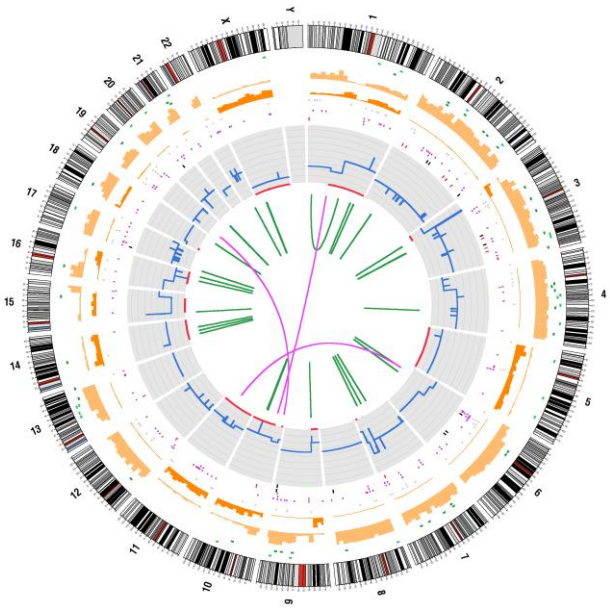
Whole genome shotgun

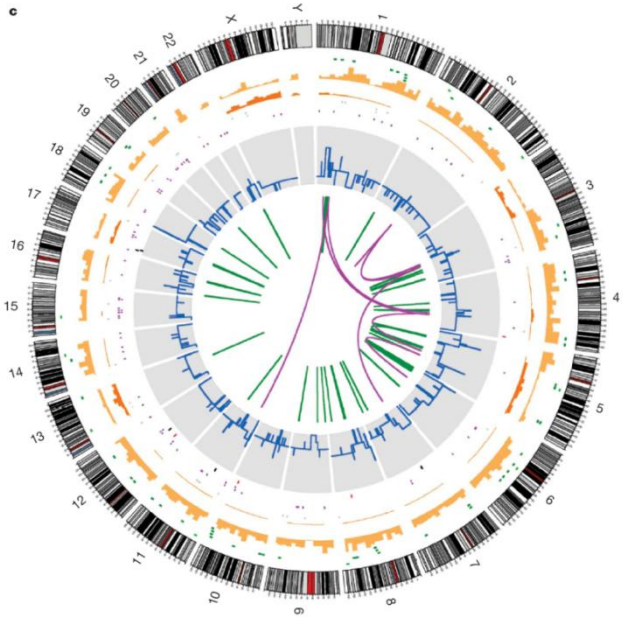
Comprehensive catalogues of somatic mutations in cancer

- Detection of all classes of somatic variant
  - base substitutions
  - insertions and deletions
  - rearrangements
  - copy number
- Detection of somatic variants in all genomic regions
  - coding exons
  - noncoding exons
  - introns
  - intergenic regions

Somatic mutations in Colo-829 and NCI-H209

	Colo-829	NCI-H209
Total Somatic Substitutions	33,345	22,910
Insertion/deletion	66	65
Rearrrangements	37	58
CODING		
missense	168	92
silent	104	36
nonsense	14	4





International Cancer Genome Consortium

- **Brain Cancer** — United States
- **Breast Cancer** — European Union / United Kingdom
- **Breast Cancer** — France
- **Breast Cancer** — United Kingdom
- **Chronic Lymphocytic Leukemia** — Spain
- **Colon Cancer** — United States
- **Gastric Cancer** — China
- **Leukemia** — United States
- **Liver Cancer** — France
- **Liver Cancer** — Japan
- **Lung Cancer** — United States

[Show]

ICGC Goal: To obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe.

International network of cancer genome projects. Nature **464**, 993-998 (15 April 2010)

HTML

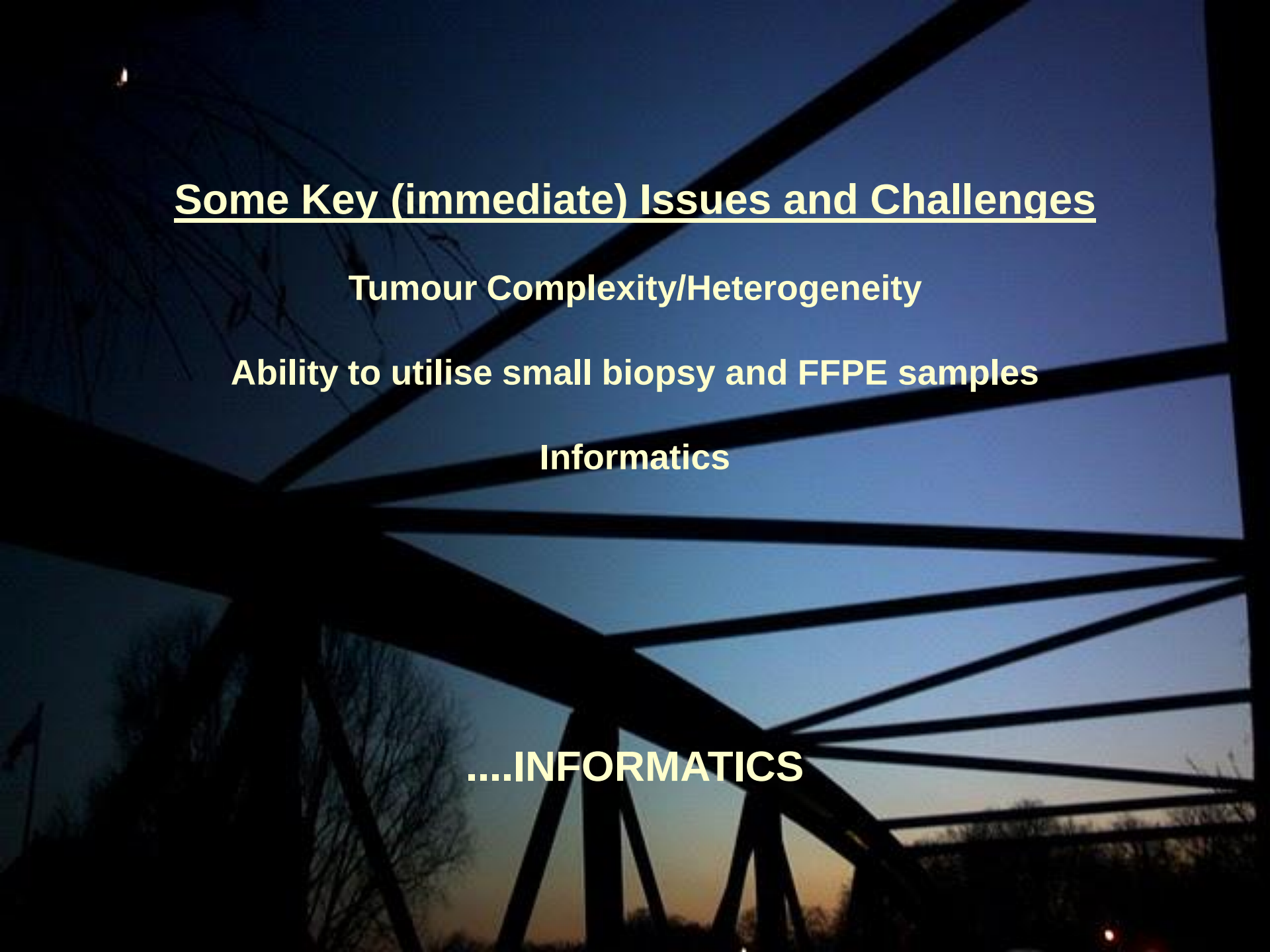
ICGC Public Presentation April 15, 2010 : PDF | PPT

International Cancer Genome Consortium (ICGC) Goals, Structure, Policies and Guidelines : HTML | PDF

Members of the ICGC

Committed Projects to date: 23

- **Lung Cancer** — United States
- **Malignant Lymphoma** — Germany
- **Oral Cancer** — India
- **Ovarian Cancer** — Australia
- **Ovarian Cancer** — United States
- **Pancreatic Cancer** — Australia
- **Pancreatic Cancer** — Canada
- **Pediatric Brain Tumors** — Germany
- **Prostate Cancer** — Canada
- **Prostate Cancer** — Germany
- **Rare Pancreatic Tumors** — Italy
- **Renal Cancer** — European Union / France



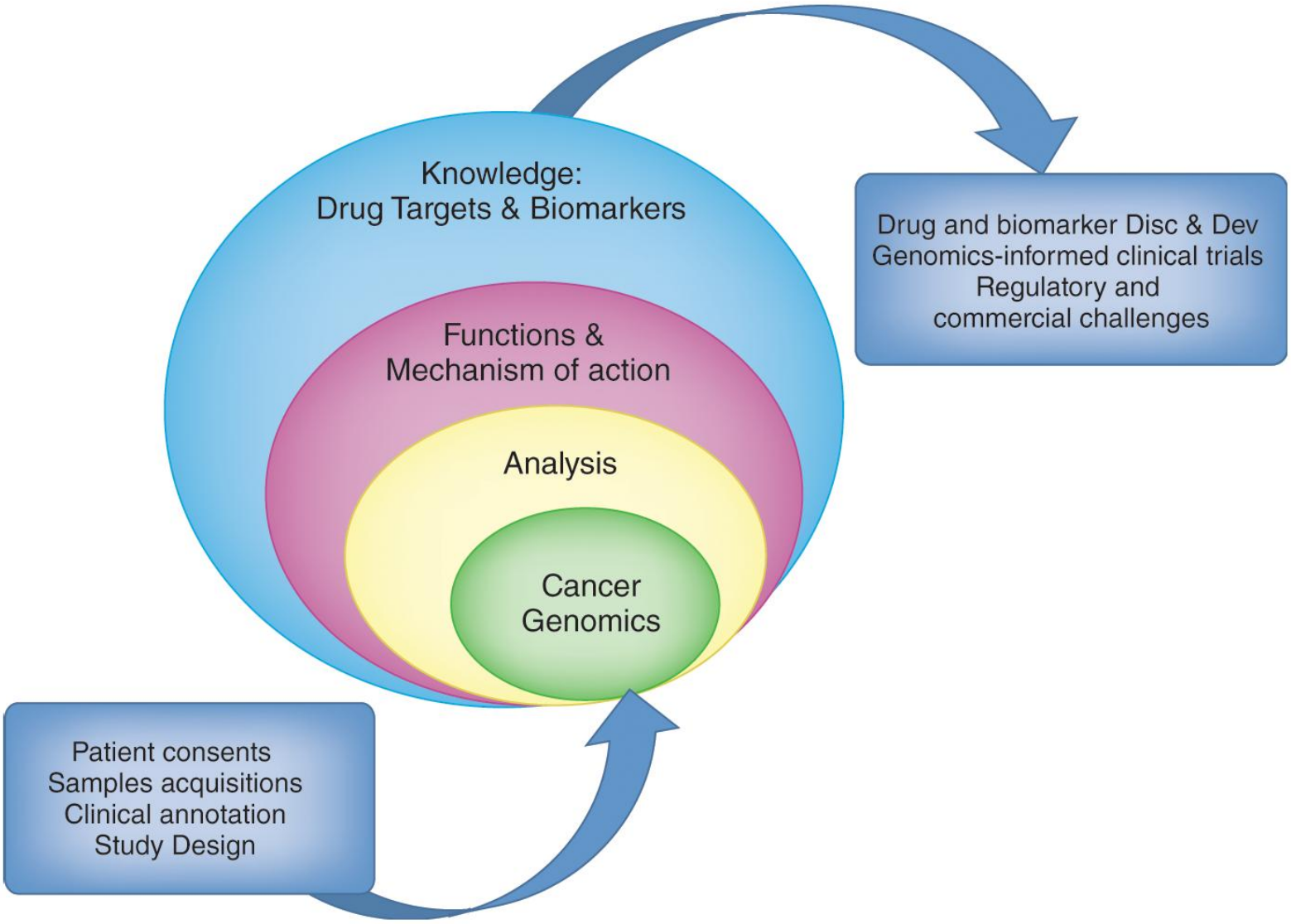
Some Key (immediate) Issues and Challenges

Tumour Complexity/Heterogeneity

Ability to utilise small biopsy and FFPE samples

Informatics

....INFORMATICS



Chin, Andersen, Futreal – Nature Medicine, in press

Cancer Genome Project

Sally Bamford
David Beare
Graham Bignell
Nidhi Bindal
Adam Butler
Helen Davies
Charlotte Dunham
Simon Forbes
Christopher Greenman
Claire Hardy
Mingming Jia
David Jones
Chai Yin Kok
Calli Latimer
King Wai Lau
Kenric Leung
Meng-lay Lin
Mark Maddison
John Marshall
David McBride
Stuart McLaren
Andrew Menzies
Lina Chen
Juok Cho
Danushka Galappaththige
Catherine Leroy

Laura Mudie
Keiran Raine
Rebecca Shepherd
Lucy Stebbings
Philip Stephens
Patrick Tarpey
Jonathan Teague
Tony Webb
Stacey Price
Ignacio Varela

Mathew Garnett
Anne McLaren-Douglas
Andrew Barthorpe
Patrick Brien
Laura Hirst
Frances Jewitt
Tatiana Mironenko
Jorge Soares
Ian Thompson

Serena Nik-Zainal
Elizabeth Murchison
Jennifer Yen

Sancha Martin
Angela Macharia
Wendy McLaughlin

Erin Pleasance
Gillian Dalgliesh

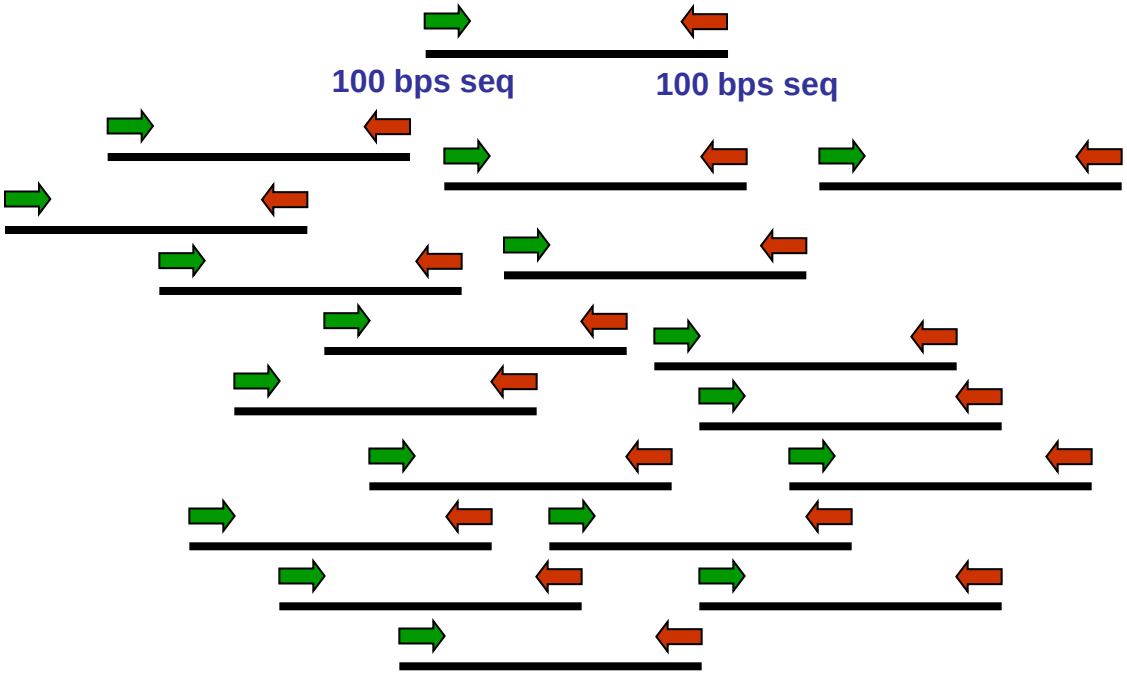
Mike Stratton, Peter Campbell, Ultan McDermott

ccRCC/PBRM1 Studies
Bin Teh
Dachuan Huang
Choon Kiat Ong
Waraporn Chan-on
Chutima Subimerb
Kyle Furge
Karl Dykema

ICGC Breast Cancer Working Group

Paired-end Reads

Random 400bp fragments of matching cancer and normal genomes
Hundreds of millions of individual molecules sequenced simultaneously



Map paired sequences back to reference genome
Compare and look for tumour-specific variants

Assay design