Sequencing Cancer Genomes and Transcriptomes

Dr. Richard Moore

Sequencing Group Leader

Genome Sciences Centre, BC Cancer Agency

Capacity at the GSC

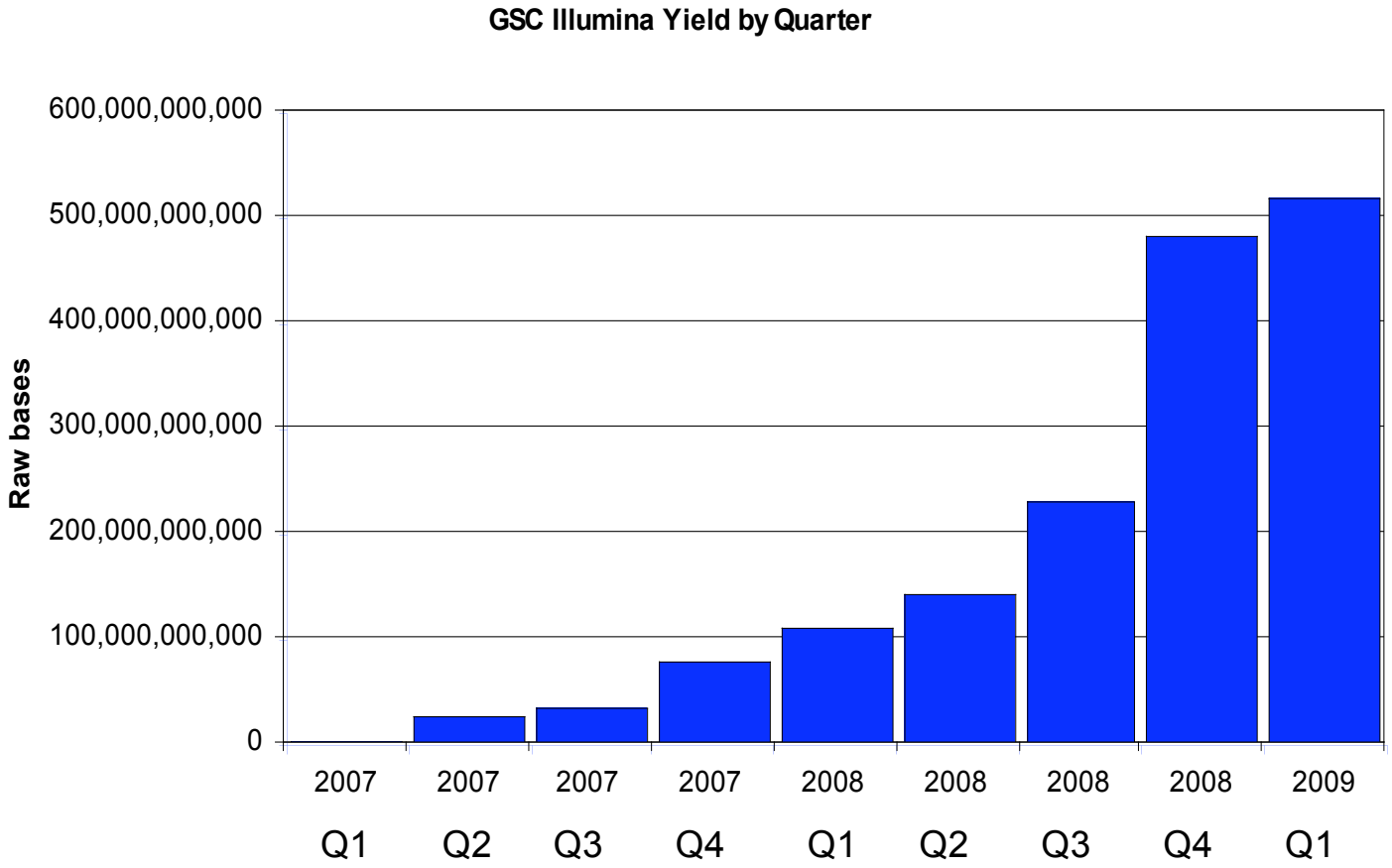
- 11 Illumina GA II
- 2 AB SOLiD
- 8 3730 xls

- Paired and Single end reads
- Production 51 bases, Tech D upto 101

- Library Types

Whole Genome Shotgun
miRNA
RNA seq / WTSS
ChIP
Bisulfite

MRE
Small RNA
MeDIP
SAGE



Total 1.6 Trillion bases

GSC Utilisation of UHTS technology

- 26 UHTS publications
  - Including Genome research and Nature Methods
- 1.6 trillion bases produced
- >1200 libraries

Examples of Current Projects

- Lobular Breast Cancer
- Adenocarcinoma - personal genomics
- Follicular Lymphoma

- Diffuse Large Bcell Lymphoma
- Ovarian Cancer
- Lung Cancer
- Acute Lymophoblastic Leukaemia
- Neuroblastoma
- Prostate Cancer
- Oligodendrioma
- Colo-rectal Carcinoma

Cancer prevalence

Life time probability of:

	developing		***dying***	
	%	one in:	%	one in:
Females	39.3	2.5	24.1	4.2
Males	44.5	2.2	28.5	3.5

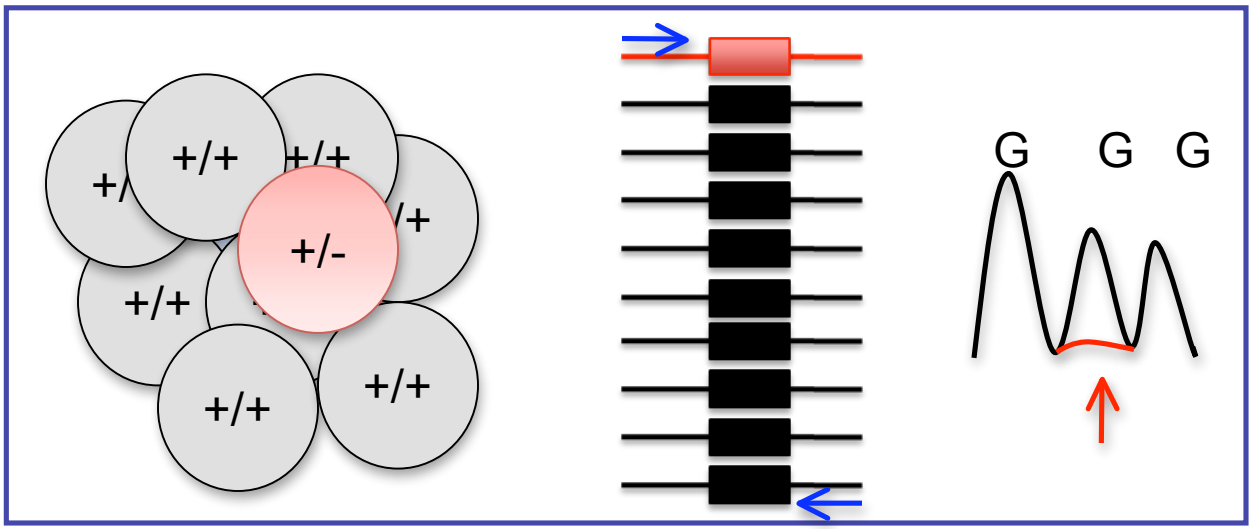
Canadian Cancer Society Statistics, 2008

Why sequence cancer genomes and transcriptomes?

- To identify cancer driver mutations and pathways.
  - Challenges: passengers vs. drivers, undiscovered SNPs, changes selected for in culture, genetic heterogeneity of cell populations…
- To identify targets for development of new therapies.
- To improve diagnostic precision and prognostic accuracy.
  - E.g. “breast cancer” describes several diseases.
- To match patients to treatments.
  - Optimize treatment modalities based on individual genes and genomes
- To understand differences in treatment response.
  - Outright failure
  - Remission / relapse (treatment resistance).
- To manage treatment failure.
  - Alternative existing therapies?

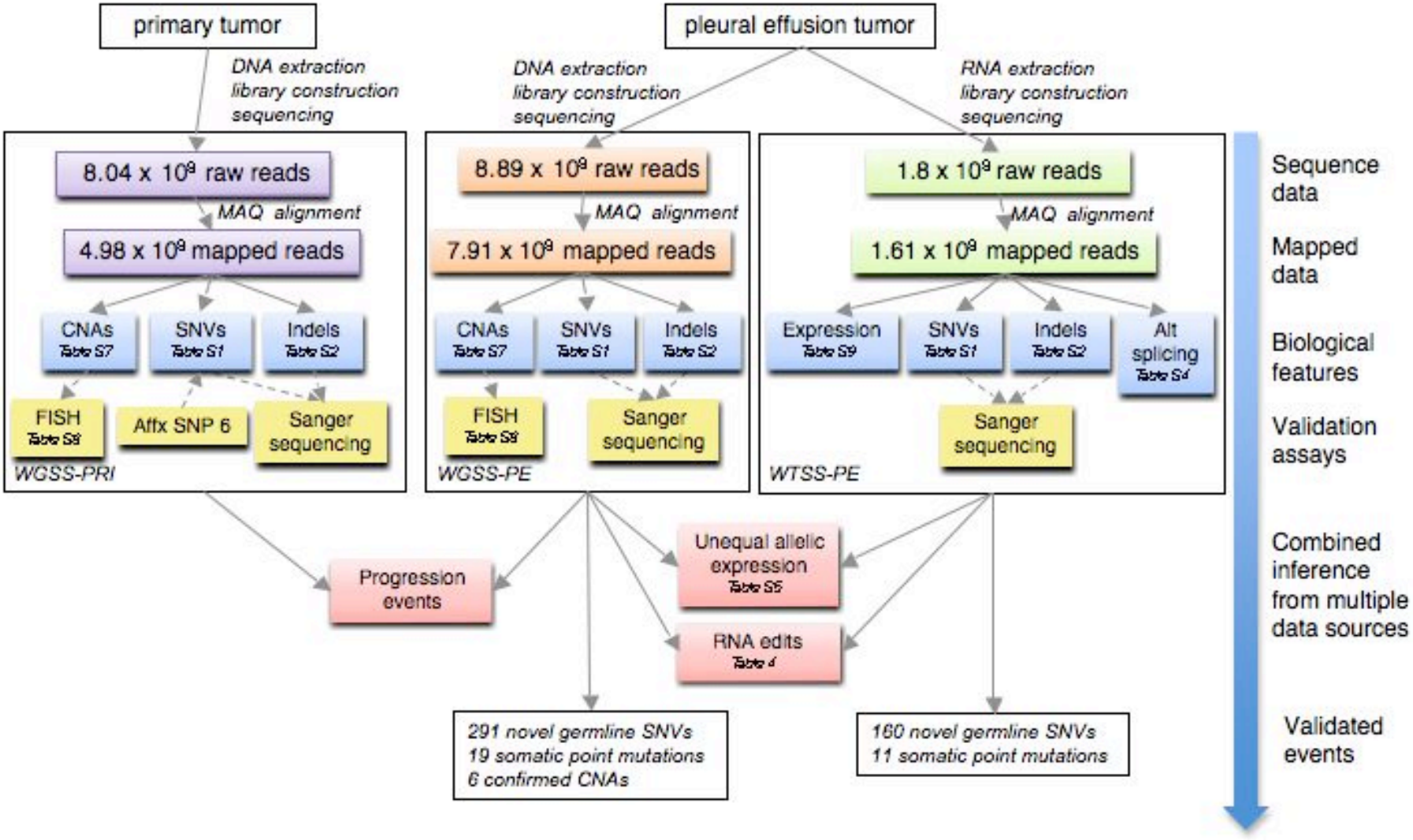
Next-generation sequencing of cancer patient samples

- Depth-dependent sensitivity to rare events; e.g. genomic changes in cellular sub-populations (tumor initiating cells?).
  - Enabled by massively redundant “single molecule” sequencing, in contrast to technologies that integrate signal from many molecules.



Lobular Breast Cancer

Lobular Breast Cancer



Samples and libraries

	WGSS-PE	WTSS-PE	WGSS-PRI
Total number of reads	889,392,298	182,532,650	804,148,860
Total nucleotides (Gb)	34.419	7.108	35.511
Number of aligned reads	790,665,100	160,919,484	497,521,910
Aligned nucleotides (Gb)	30.599	6.266	21.971
Estimated error rate	0.015	0.013	0.019
Estimated depth (non-gap regions)	9.809	N/A	6.776
Canonically aligned reads	721,929,588	109,093,616	439,409,649
Percent reads aligned canonically	91.31	67.79	88.32
Unaligned reads	98,727,198	21,613,166	306,626,950
Mean read length (bp)	38.7	38.94	44.16

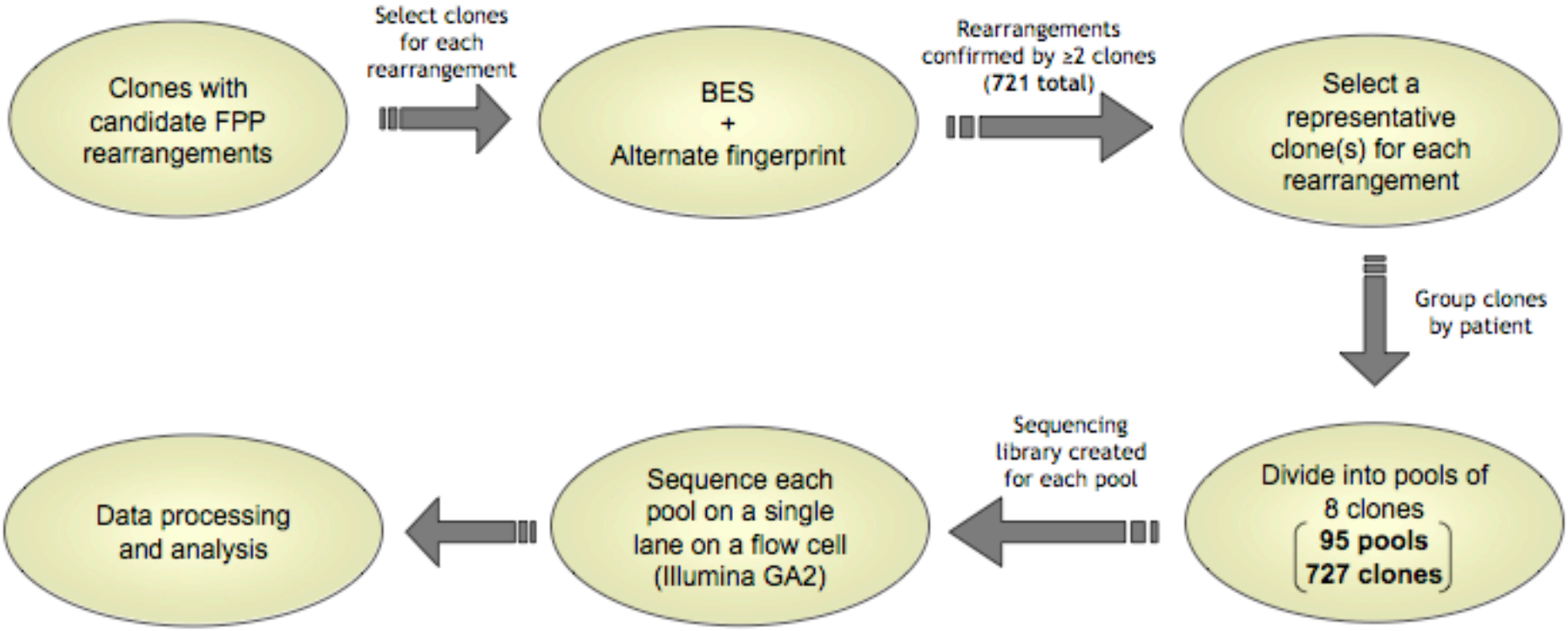
Summary

- Alternate spicing shown in the estrogen signalling pathway
- 25 NS somatic mutations
- 16 only in the metastaic tumour
- 2 RNA editing somatic NS mutations
- Novel amplicon region in the Insulin receptor

Follicular Lymphoma

BAC re-sequencing of large scale rearrangements

Rearrangement confirmation and clone sequencing pipeline



- 691 of the 721 total rearrangements were sampled
- All 253 distinct rearrangements were sampled in at least one patient

  ⇒ Not all recurrent rearrangements sampled in all patients

Summary of validated large-scale somatic events

We have identified and validated 52 somatic and 66 germline breakpoints, corresponding to 38 and 41 distinct events. Examples of somatic mutations include:

Event type	Chromosome band(s)	Event size (kb)	Patient ID	Junction feature	Affected gene(s)
DEL	1p22.3	558	8	Microhomology	LPAR3, MCOLN2, MCOLN3, WDR63, SYDE2, C1orf52, BAG and last exon of BCL10
DEL	9p24.1	251	10	Blunt	PTPRD
DEL	9p21.3	124	16	Microhomology	CDKN2A
DEL	2p11.2	193-412	8, 13, 16, 17	All features	AbParts, IGK
DEL	22q11.22	112-861	6, 17	Sequence additions	AbParts, IGL
DEL	14q32.33	395	19	Sequence additions	AbParts, IGH
TRX	t(14;18)(q32;q21)	NA	6, 8, 10, 13, 16	Sequence additions	IGH and BCL2
TRX	t(10;12)(q25.1;q23.1)	NA	10	Microhomology	SORCS1
TRX	t(6;7)(q15;q36.1)	NA	20	Microhomology	CDK5
TRX	t(1;3)(q23.3;p24.3)	NA	21	Blunt	RFTN1
INV	3q27.3	718/722	14, 21	All features	BCL6 and ST6GAL1
INV	6q25.2	80	21	Microhomology	OPRM1
DUP	10q22.2-10q23.32	18,126	19	Sequence additions	AP3M1 and BTAF1
RNG	4q12	137	20	Sequence additions	KIT

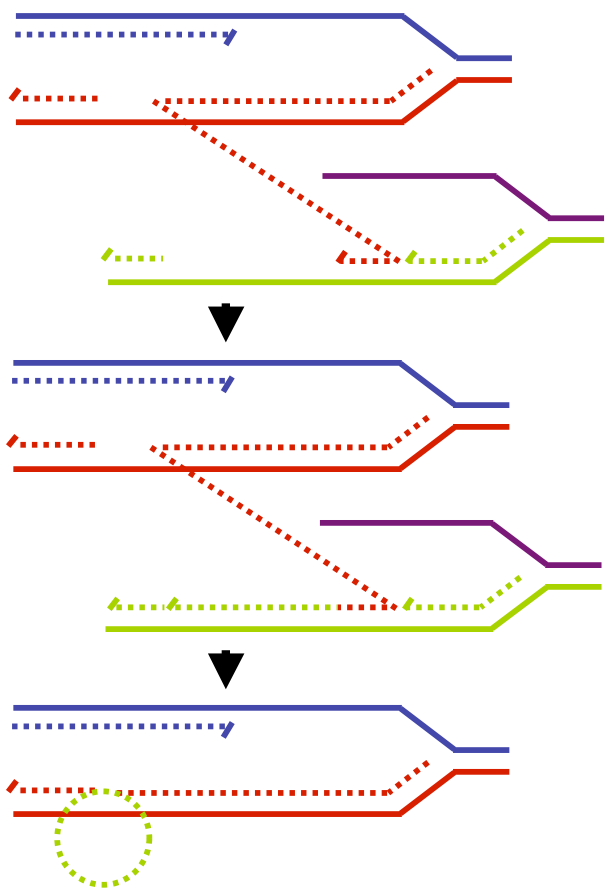
DEL, Deletion; TRX, Translocation; INV, Inversion; DUP, Duplication; RNG, Complex rearrangement; NA, Not applicable

Fork Stalling and Template Switching – a model for nucleotide additions at breakpoint junctions

A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders

Jennifer A. Lee,[1] Claudia M.B. Carvalho,[1] and James R. Lupski[1,2,3,*]
[1]Department of Molecular and Human Genetics
[2]Department of Pediatrics
Baylor College of Medicine, Houston, TX, 77030, USA
[3]Texas Children's Hospital, Houston, TX, 77030, USA
*Correspondence: jlupski@bcm.tmc.edu
DOI 10.1016/j.cell.2007.11.037

Cell. 2007 Dec 28;131(7):1235-47



Concluding remarks

- To date we have identified 38 distinct somatic events from 20 primary follicular lymphoma patient tumour samples.
- 40% of the rearrangement events assayed are acquired in the tumour samples.
- Having sequence level resolution of breakpoints is informing on mechanistic insights.
- For the first time in cancer, to our knowledge, we have demonstrated that sequence additions at rearrangement breakpoints are consistent with the FoSTeS model of DNA replication.

Personal cancer genomics:

Adenocarcinoma, pulmonary metastases -

non-responsive to treatment (erlotnib)

Questions

- In this case (rare tumor, no standard chemotherapy options), can next generation genome analysis provide clues as to the expression and mutational profiles of known drug targets?
- Can next generation genome analysis provide insight into the apparent resistance of the tumor to erlotinib, despite apparent amplification of EGFR?

Samples and libraries

	Type	Source	Amount	Total reads	% aligned to genome	Sequence yield total (base-pairs)	Sequence yield aligned (base-pairs)
	RNA-seq	Normal: Blood	10ug Total RNA	85,795,100	72.9	3,603,394,200	2,625,754,824
	Genome Shotgun	Normal: Blood	10ug gDNA	40,793,289	83.7	1,713,318,138	1,434,873,300
	RNA-seq lite (cDNA amp)	Cancer: Lung Biopsy	10ng Total RNA	27,897,5099	69.2	11,716,954,158	8,104,150,782
	RNA-seq lite (cDNA amp)	Cancer: Lung Biopsy	10ng Total RNA	431,797,605	69.7	18,135,499,410	12,645,184,944
	RNA-seq lite (RNA amp)	Cancer: Lung Biopsy	2ng Total RNA	6,328,326	66.3	265,789,692	176,282,652
	Genome Shotgun	Cancer: Lymph Node FFPE	10ug gDNA	2,942,711,791	87.8	123,593,895,222	108,551,254,728
Totals:				3,786,401,210		159,028,850,820	133,537,501,230 ~84%

Data analysis

- Transcriptome libraries:
  - Digital gene expression profiling, comparing tumor read counts across ~23 in-house WTSS libraries.
  - Mutation detection.
- Genome libraries:
  - Digital karyotyping (copy number analysis), comparing tumor genome sequence to PB ("normal") genome sequence and the Yoruban sequence.
  - Mutation detection.
- Integration with drug bank:
  - Relate mutations, copy number alterations, and gene expression data to known drugs.

Drug Bank

- www.drugbank.ca
- 4,408 drugs in drugbank for 4535 targets.
- 1,359 FDA-approved drugs for 1613 targets
- 189 "cancer" drugs for 291 targets.

Intersection of drug bank with genomic features and cancer pathways

- 30 amplified genes in cancer pathways.
- 76 deleted genes in cancer pathways.
- ~400 up- and ~400 down- regulated genes.
  - 19 of these in known cancer pathways
- Single base changes (transcriptome):
  - 303 non-synonymous point mutations in 177 genes
  - 4 genes in cancer pathways
- Other:
  - 233 novel coding SNPs,
  - 25 genes affected by indels
  - 126 genomic deletions spanning exons for 83 genes
  - No intersections with known cancer pathways

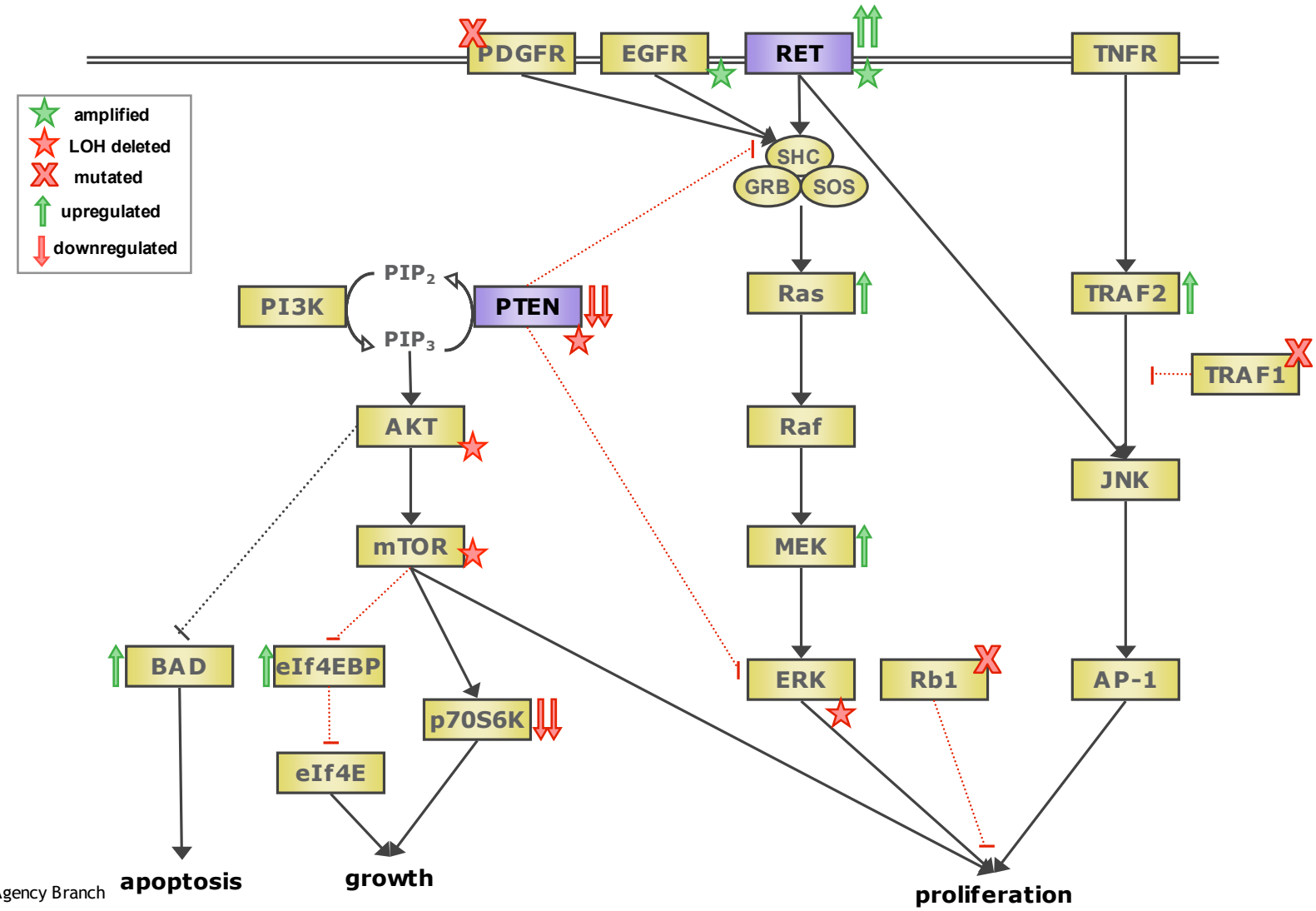
Proposed therapeutic interventions

Genomic aberration	Drug	Indication/Known mechanism
RET is amplified & upregulated, RAS is upregulated. MEK is upregulated. PDGFRA is mutated. RET-RAS-RAF-MEK pathway is further activated by PTEN loss.	Sunitinib	Targets RET/PDGFRs/VEGFR/KIT/CSF1R/FLT3 for GIST, RCC, and thyroid cancer.
	Sorafenib	Targets RET/BRAF/VEGFR2&3/PDGFRB/KIT/FLT3 for advanced RCC, HCC and thyroid cancer.
	Vandetanib	Targets VEGFR/EGFR for NSCLC.
	Motesanib	Targets VEGFR/PDGFR/KIT/RET for thyroid cancer.
	Sulindac	An NSAID thought to target COX-1 and COX-2 for inflammation but can inhibit HRAS.
PTEN downregulation (heterozygous deletion) and RB1 truncation. PTEN loss also increases PI3K-AKT-mTOR pathway activity.	Rosiglitazone +carboplatin	Activates PPAR-gamma, produces effects similar to inhibiting EGFR, by downregulating Akt and mTor and upregulating PTEN.
	rapamycin	mTOR inhibitor. An approved immuno-suppressant use to prevent rejection in organ transplants.
ROCK1 is the only gene in a very highly amplified region.	fasudil	Rho-kinase (ROCK1) inhibitor. Inhibits VEGF-induced angiogenesis. However, ROCK1 can activate PTEN, so should not use with PTEN-upregulating strategies.

Sunitinib was chosen because it targets the most number of expressed genes in the tumour. It inhibits RET, PDGFRA, CSF1R.

Sorafenib is also able to inhibit a mutant RET and B-RAF, and was thus selected as secondary treatment in case the patient did not respond or acquired resistance to sunitinib.

Integrating pathways, mutations and gene expression



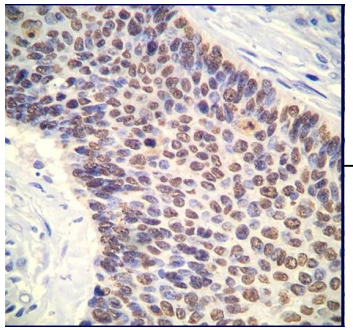
Response to Sunitinib

After 29days on Eroltinib tumours increased in size by >20%

After 28 days on Sunitinib tumour reduced in size by >20%

Infectious Agents and Cancer

Genomic Subtraction For Infectious Agent Detection



Tumour Biopsy

Total RNA extraction

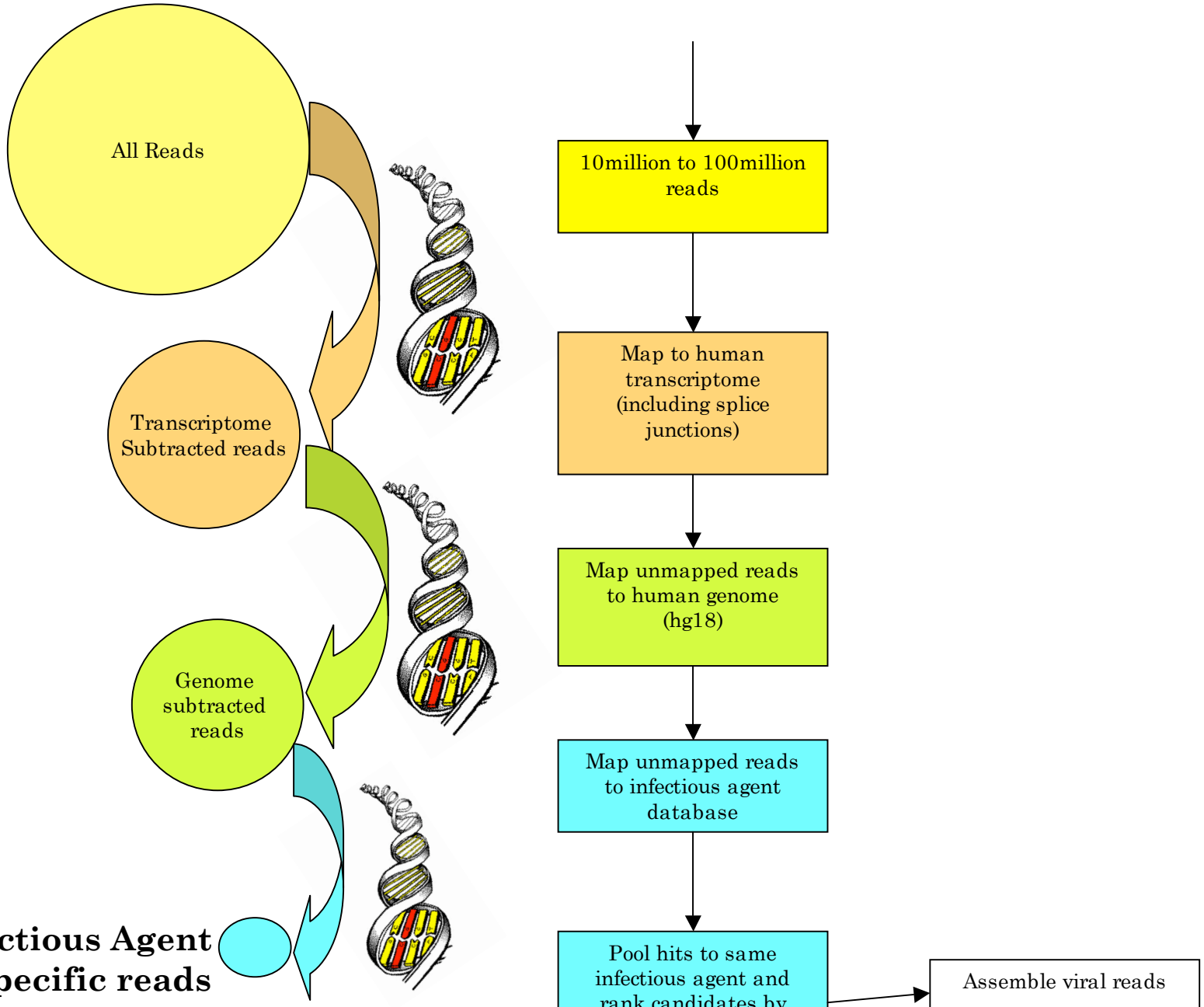
PolyA selection

Or

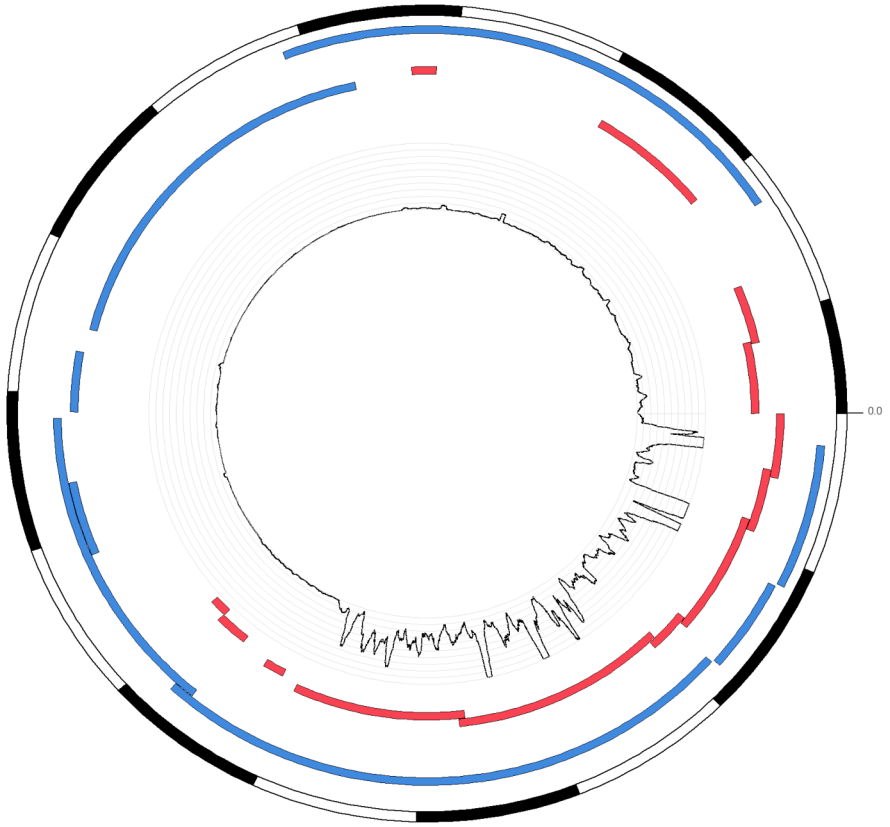
Ribominus depletion

cDNA fragment library construction 150-300bp

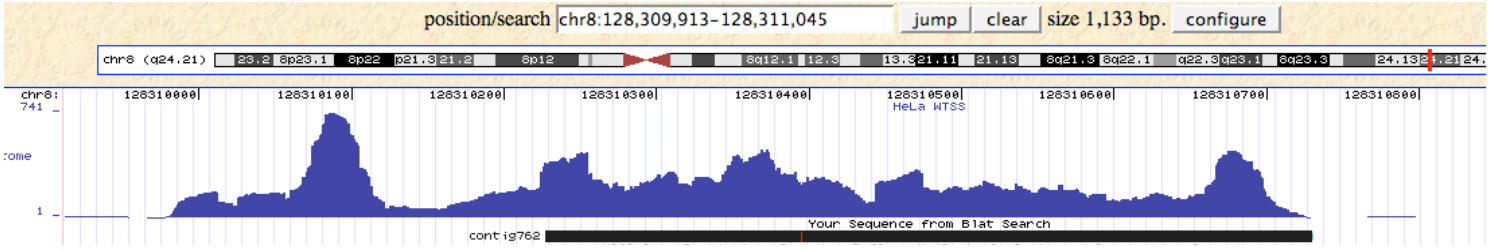
Sequencing on Illumina GA



De-novo assembly of unmapped reads reveals viral RNA and integration site



- HeLa is derived from a cervical cancer positive for HPV infection
- De-novo assembly of unmapped reads reveals HPV-derived sequences
- A few HPV contigs contain human sequence (reveal integration site of virus)



Acknowledgements

Bilenky, Misha
Birol, Inanc
Chuah, Eric
Corbett, Richard
Griffith, Obi
Hirst, Martin & team
Holt, Rob
Jones, Steve
Li, Yvonne
Marra, Marco
Moore, Richard & team
Morin, Ryan
Morozova, Olena
Pugh, Trevor
Roscoe, Robyn & team
Shah, Sohrab
Thiessen, Nina
Varhol, Richard & team
Zhao, Yongjun &team

Cancer patients and research advocates

Aparicio, Sam
Huntsman, David
Jones, Steven
Laskin, Janessa

Funding:
Genome Canada
Genome British Columbia
BC Cancer foundation
CIHR
NCI (USA)
NCI (Canada)
Western Economic Diversification