



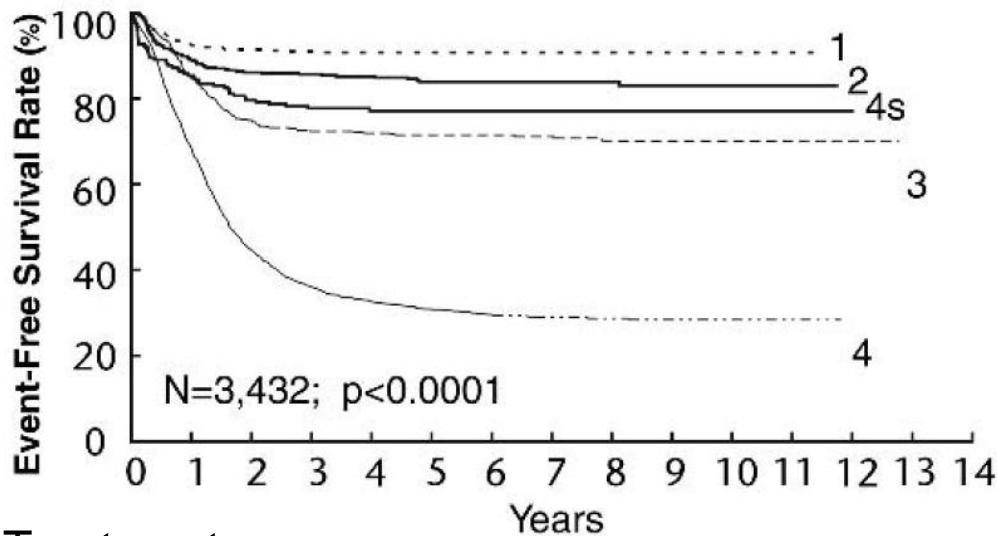
A wide spectrum of somatic mutations in high-risk neuroblastoma

Olena Morozova
Marra Lab
Genome Sciences Centre

VanBUG
October 13, 2011

Neuroblastoma (NB): a childhood cancer of the sympathetic nervous system

- International Neuroblastoma Staging System (INSS) Stage 4 disease
- Dissemination to bone, bone marrow, liver, skin, other organs
- 40% of all NB cases
- About half of high-risk cases have *MYCN* oncogene amplification

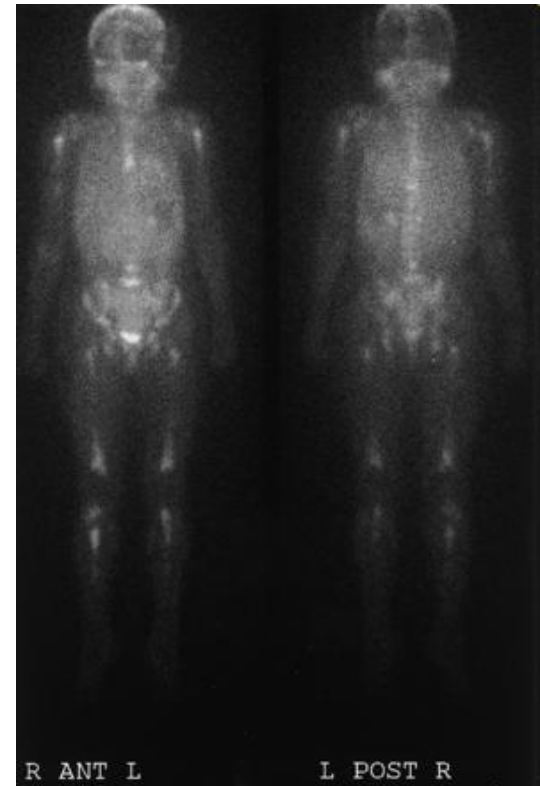


Treatment:

Chemotherapy, surgery, radiation, autologous stem cell transplant, biological therapy, immunotherapy

Prognosis:

< 40% cure; no known cure after relapse



¹²³I-MIBG scan
Kushner, 2004

Neuroblastoma TARGET initiative



Therapeutically Applicable Research to Generate Effective Treatments (TARGET) Initiative (NCI)

Aim: To identify new therapeutic targets for pediatric cancers using state-of-the-art genomic approaches (pediatric counterpart to The Cancer Genome Atlas, TCGA)

Neuroblastoma TARGET (projects completed to date):

Affymetric Human Exon 1.0 ST Array (>200 cases)

Illumina 550K SNP chip (>200 cases)

Sanger re-sequencing of 118 candidate genes (~200 cases)

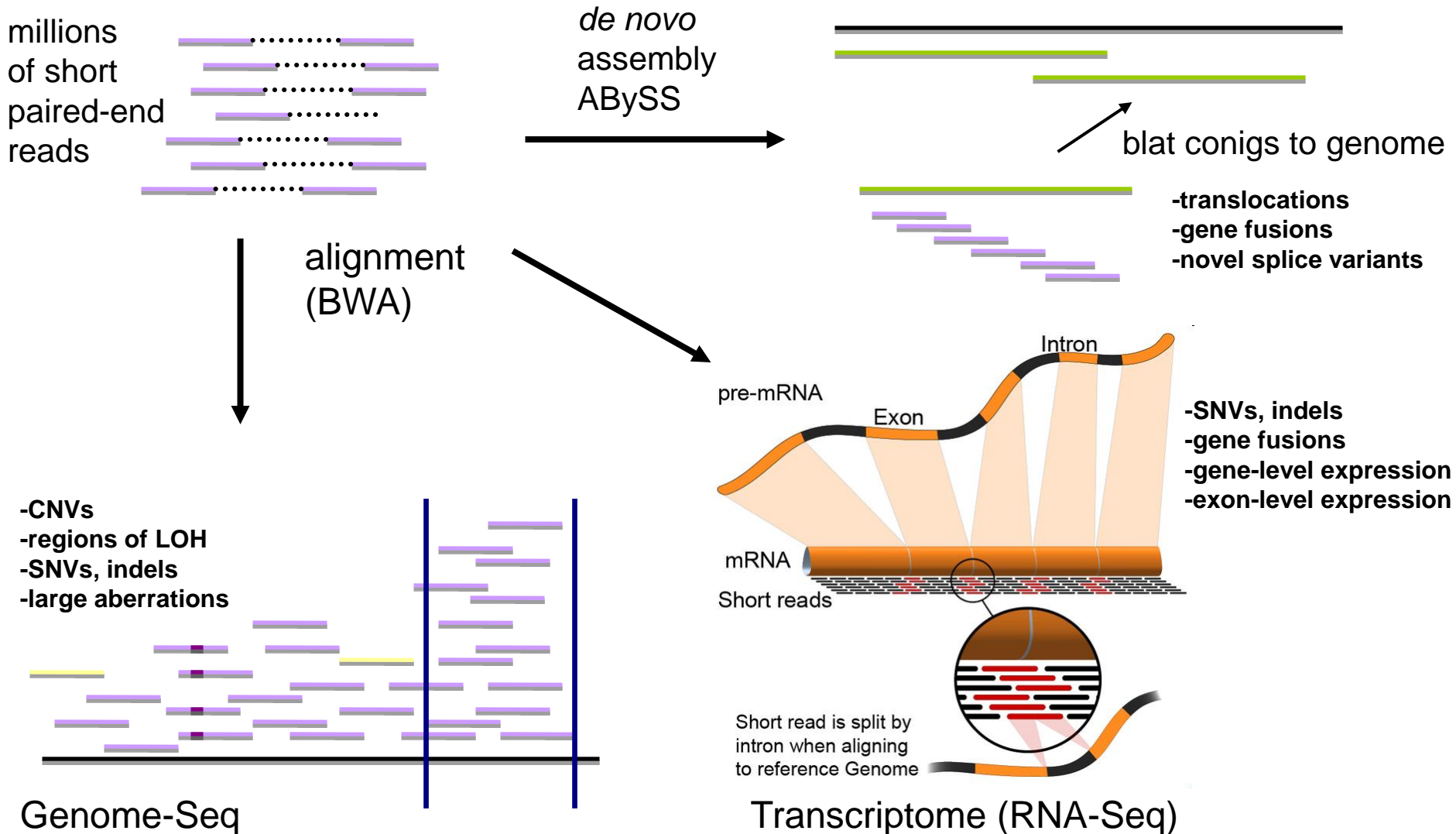
-- Discovery of somatic ALK mutations in ~7% of neuroblastomas (most frequently mutated gene)

-- Was the wrong candidate gene list selected?

Next-generation sequencing (NGS) of whole genomes, exomes and transcriptomes

(99 cases)

GSC: whole genome and transcriptome sequencing of 10 NB cases

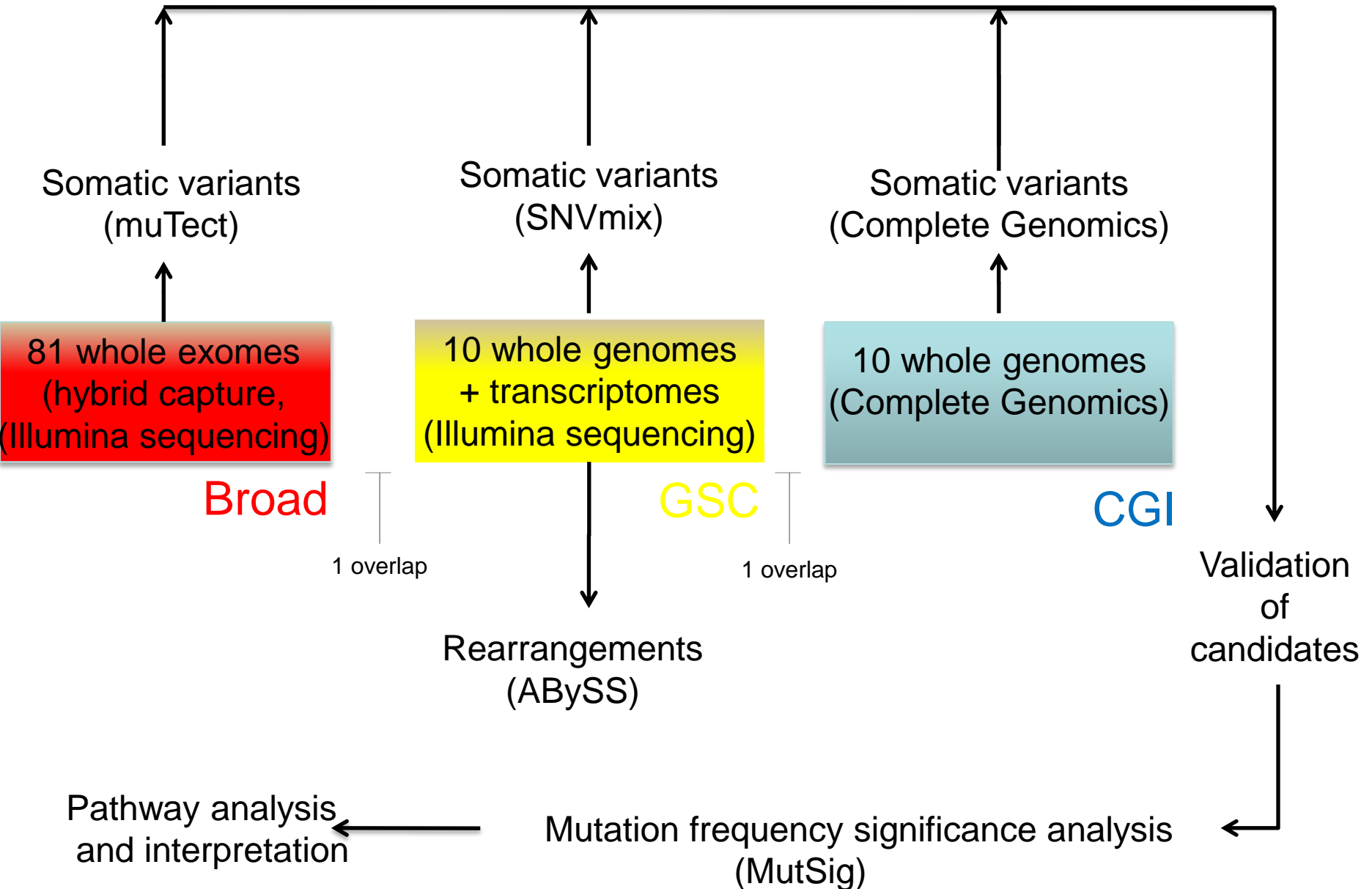




Spectrum of somatic point mutations in 10 NBL genomes

Case	Coding	UTR, splice site	RNA-Seq support (coding and UTR)	intron	Total	Notable genes with coding/UTR change
NBL1	2	0	0	430	1181	ALS2CR8
NBL2	4	20	13	637	2086	DIDO1, SOX5, EVC
NBL3	5	8	6	574	1744	CYP8B1, MUC20
NBL4	9	13	14	992	3170	RXR8, CHST11, CDC25A
NBL5	19	23	16	852	2335	CREBBP, CDKN1A, SPECC1, STAT5B
NBL6	14	22	19	790	2352	FANCD2, SMAD2
NBL7	3	9	7	552	1580	ATM, PLXNC1, EVC
NBL8	17	9	18	1208	3736	SPECC1, DIDO1, ABL2, MLL5, DISC1, germline ALK
NBL9	9	10	13	741	2289	MSR1, BEST4, URG4
NBL10	6	5	7	574	1806	TERF1, CNTN2, CYP2F1

Neuroblastoma TARGET next-generation sequencing initiatives



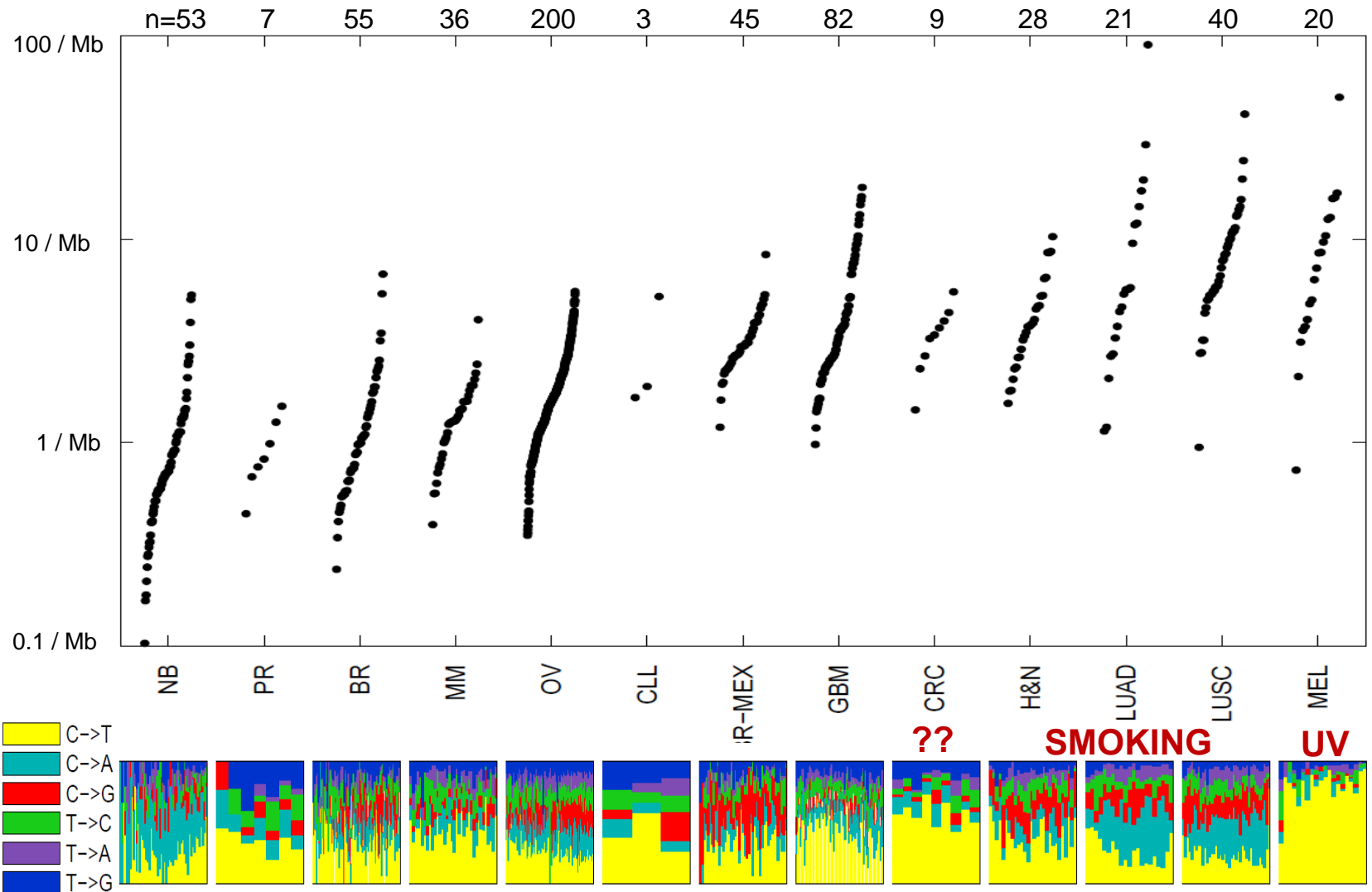
Integrating data from three different NGS approaches

- Three different sequencing approaches
 - Illumina exome capture (151X of target bases)
 - Illumina whole genome sequencing (30X of mappable genome)
 - CGI whole genome (60X of mappable genome)
- How to compare mutations and mutation frequencies
 - Restrict analysis to coding space as in the Broad exome
 - Need to correct for “callable bases” in each approach
 - CGI: use company’s definition of “call” and “no call” (require “call” at both alleles)
 - Illumina: use read evidence
 - Positions with 10 read coverage in tumor and 5 read coverage in normal

Integrating data from three different NGS approaches: importance of controls

- Cases sequenced by two platforms
 - Broad/GSC common case
 - Somatic non-silent mutation rate based on the Broad data 0.59
 - Somatic non-silent mutation rate based on the GSC data 0.65
 - CGI/GSC common case
 - Somatic non-silent mutation rate based on the CGI data 0.58
 - Somatic non-silent mutation rate based on the GSC data 0.66
- Global control: cases analyzed by different platforms should not stand out

Neuroblastoma has the lowest somatic mutation rate across cancers



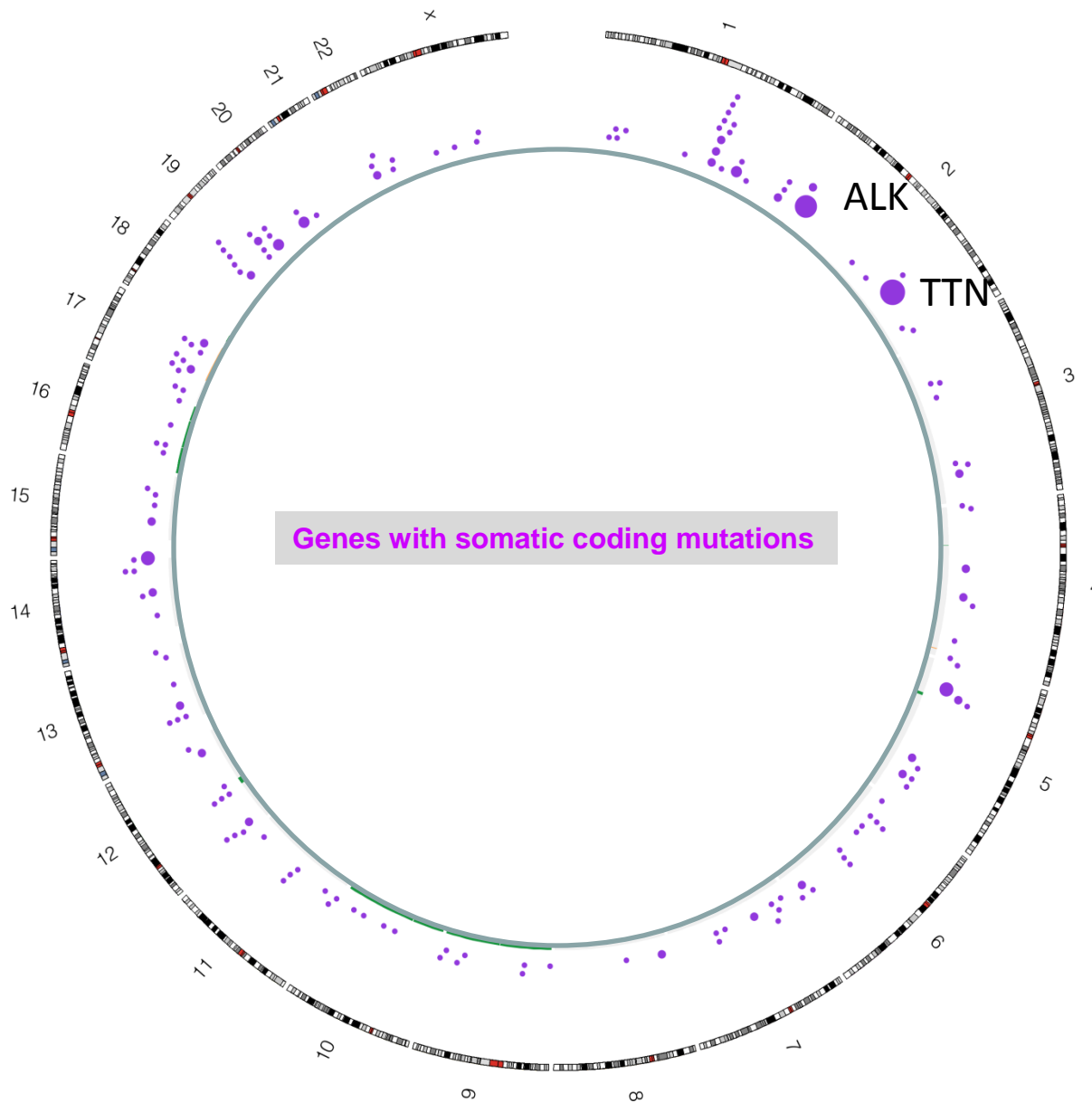
Courtesy of Gad Getz

MutSig: significance analysis of mutations

- For each gene test null hypothesis that observed mutations in that gene are a consequence of random background mutation
- Correct for callable bases, the length and composition of the gene and background mutation rates in different sequence contexts

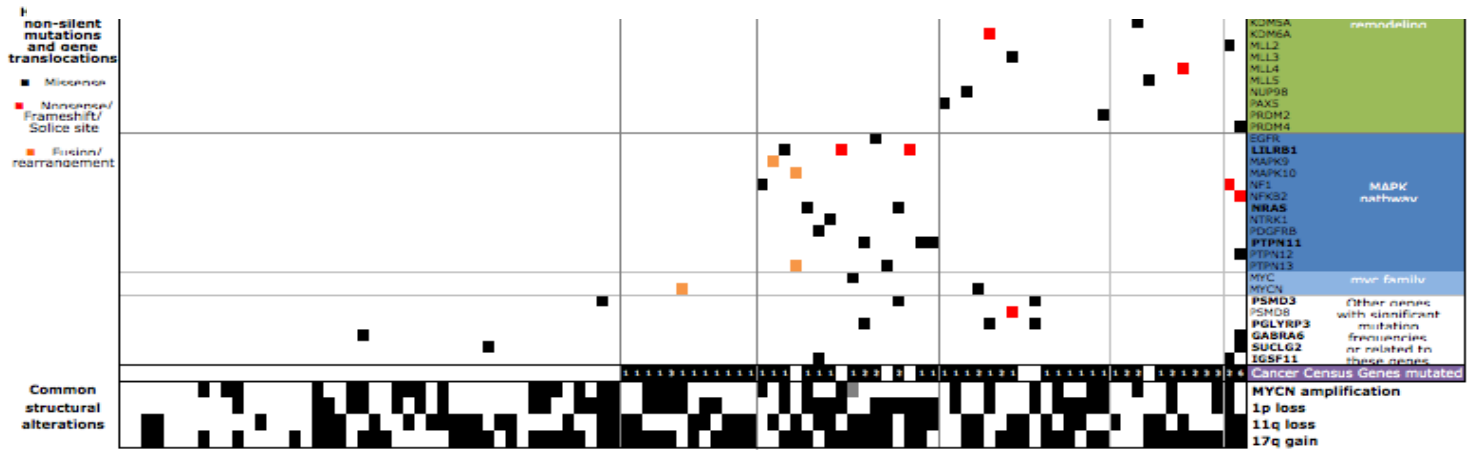
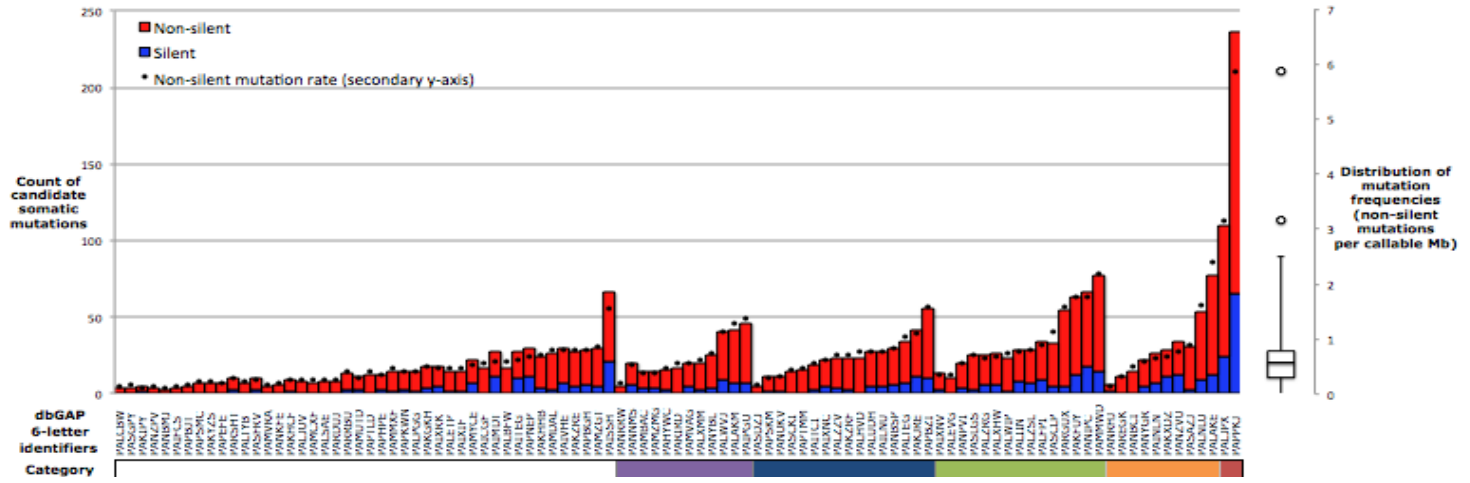
Gene	Description	Mutations	Patients	Unique sites	q-value no HM	q-value with HM	Expressed in 10 neuroblastoma transcriptomes
ALK	anaplastic lymphoma receptor tyrosine kinase	9	9	6	7.7×10^{-7}	2.6×10^{-6}	Yes
IKZF3	IKAROS family zinc finger 3 (Aiolos)	4	3	4	0.0047	0.0077	Yes
PSMD3	proteasome 26S subunit, non-ATPase, 3	3	3	2	0.0069	0.011	Yes
PGLYRP3	peptidoglycan recognition protein 3	3	3	3	0.045	0.065	No
LILRB1	leukocyte immunoglobulin-like receptor, subfamily B, member 1	3	3	3	0.071	0.085	Yes
PTPN11	protein tyrosine phosphatase, non-receptor type 11	3	3	3	0.13	0.17	Yes
NRAS	neuroblastoma RAS viral (v-ras) oncogene homolog	2	2	2	0.17	0.17	Yes
GABRA6	gamma-aminobutyric acid (GABA) A receptor, alpha 6	2	2	2	1.00	0.17	No
SUCLG2	succinate-CoA ligase, GDP-forming, beta subunit	2	2	2	1.00	0.17	Yes
IGSF11	immunoglobulin superfamily, member 11	2	2	2	1.00	0.18	Yes

Genes with somatic mutations in 99 high-risk neuroblastoma

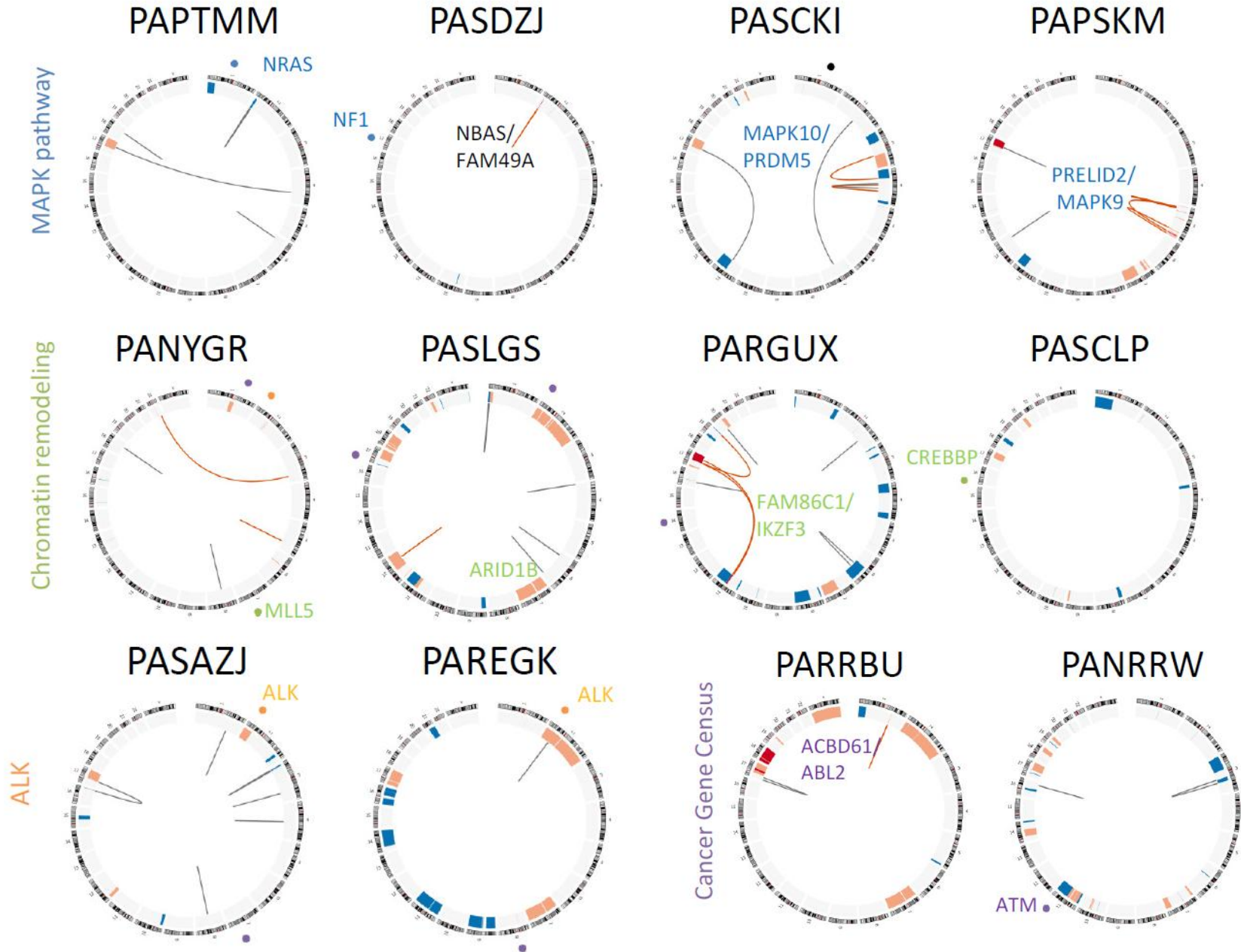


ALK is the most frequently mutated gene in NBL that passed the significance analysis (MutSig)

All known mutations in ALK verified by NGS with no FPs

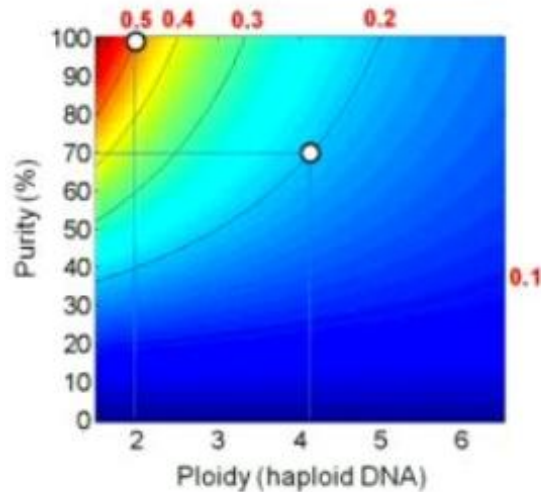


A wide spectrum of somatic mutations in NB genomes



Lessons from the NBL project

- Circulating tumor DNA found in blood complicates somatic mutation calls
 - Refine computational methods
- Tumor samples that we are sequencing are not diploid and not pure
 - Refine computational methods (CoNAn-SNV, muTect)



Purity	Ploidy	Min Allele Fraction
100%	2	.5
100%	4	.25
80%	3	.28
60%	3	.23

- Low mutation rates imply low rates of passenger events (signal indistinguishable from noise)
 - Refine methods for significance analysis of mutations
 - More samples (difficult for rare diseases)
 - Use biological knowledge to help interpret results

Biological conclusions

- Somatic mutations are relatively rare in neuroblastoma
 - *ALK* appears to be the most frequently mutated gene and may provide a tractable therapeutic target (~7% cases)
 - Mutations in RAS/MAPK pathway define a new subtype of NB (distinct from *ALK* mutants)
 - Mutations in chromatin remodeling genes in 1% cases (some overlap with *ALK* mutants)
 - 3% of high-risk cases display a hypermutator phenotype that can be linked to mutations in DNA repair genes (e.g. *MLH1*)
- Coding sequence analysis suggests a potential therapeutic avenue in ~30% of cases

Acknowledgements



PhD Advisor

-- Marco Marra

PhD Committee

-- Sam Aparicio
-- Angela Brooks-Wilson
-- Paul Pavlidis

Bioinformatics

-- Steven Jones
-- Inanc Birol
-- Martin Krzywinski
-- **Richard Corbett**
-- Mikhail Bilenky
-- Sa Li
-- Katayoon Kasaiian
-- **Karen Mungall**
-- Anthony Fejes
-- Readman Chiu
-- Shaun Jackman
-- Jenny Qian
-- Nina Thiessen
-- Johnson Pang
-- **Allan Lo**

Systems

-- Greg Stazyk
-- Roland Santos

LIMS

-- Richard Varhol
-- Ran Guin
-- Eric Chuah
-- Tina Wong

Projects

-- Cecelia Suragh
-- Robyn Roscoe
-- Karen Novik

Marra Lab

-- Ryan Morin
-- Malachi Griffith
-- Rodrigo Goya

Library Construction and Sequencing

-- Rob Holt
-- Martin Hirst
-- Richard Moore
-- Yongjun Zhao
-- Anna-Liisa Prabhu
-- Tesa Severson
-- Baljit Kamoh
-- Heather Kirk
-- Noreen Dhalla
-- Merinda Deng
-- Helen McDonald
-- Pawan Pandoh
-- Angela Tam
-- Thomas Zeng
-- Michelle Moksa
-- Diana Mah
-- Kevin Ma
-- Susanna Chan

Neuroblastoma TARGET

-- **Trevor Pugh**, Broad Institute
-- Matthew Myerson, Broad Institute
-- John Maris, Children's Hospital of Philadelphia
-- Robert Seeger, Children's Hospital Los Angeles
-- Javed Khan, National Cancer Institute
-- Edward Attiyeh, Children's Hospital of Philadelphia
-- Daniela Gerhard, National Cancer Institute
-- Malcolm Smith, National Cancer Institute
-- Michael D. Hogarty, Children's Hospital of Philadelphia
-- Shahab Asgharzadeh, Children's Hospital Los Angeles

Other

-- Neuroblastoma patients and parents



Michael Smith Foundation for
Health Research



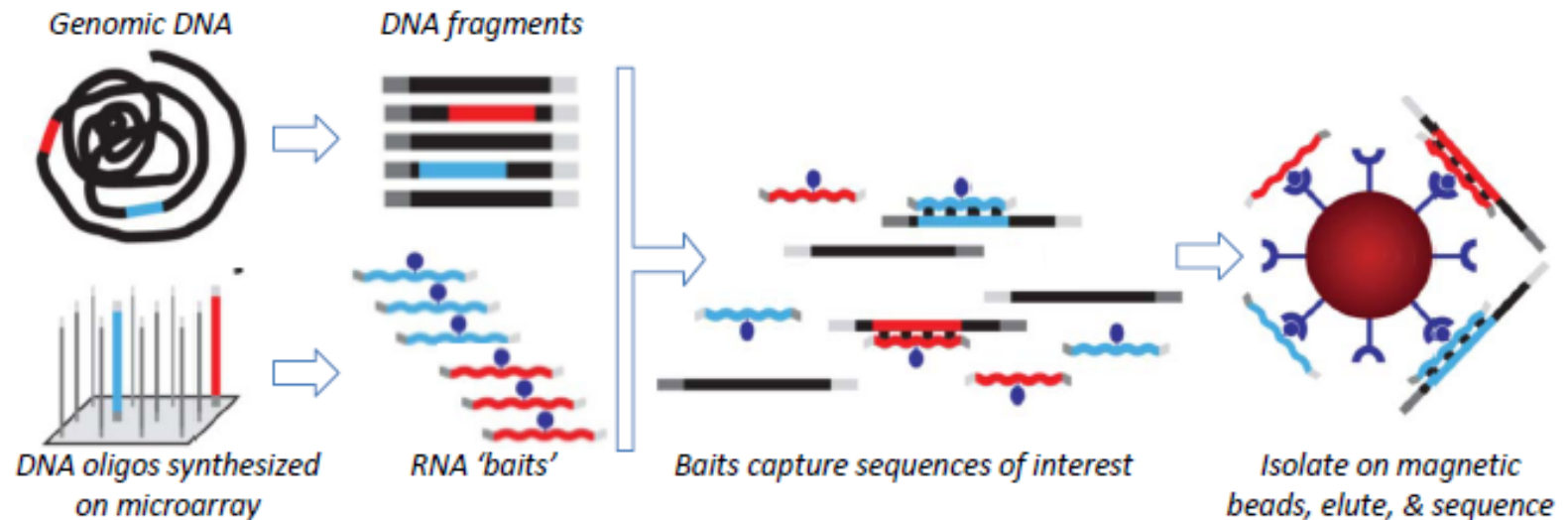
The Jordan Hopkins Foundation



Whole Exome Sequencing at the Broad

In-solution hybrid-capture using RNA 'baits'

Gnirke *et al.* Nat Biotechnol. 2009 Feb;27(2):182-9.



277,944 baits (170 bp each) targeting:

185,961 exons from 18,380 genes (CCDS + RefSeq)

45 Mbp of non-overlapping baited sequence

33 Mbp of exonic sequence