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

- de novo assembly ABySS
- blat conigs to genome
- translocations
- gene fusions
- novel splice variants
- SNVs, indels
- gene fusions
- gene-level expression
- exon-level expression

millions of short paired-end reads

alignment (BWA)

-CNVs
-regions of LOH
-SNVs, indels
-large aberrations

Genome-Seq

Transcriptome (RNA-Seq)

RNA-Seq image by Rodrigo Goya
# Spectrum of somatic point mutations in 10 NBL genomes

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

- Somatic variants (muTect)
  - 81 whole exomes (hybrid capture, Illumina sequencing)

- Somatic variants (SNVmix)
  - 10 whole genomes + transcriptomes (Illumina sequencing)

- Somatic variants (Complete Genomics)
  - 10 whole genomes (Complete Genomics)

Broad

GSC

CGI

Validation of candidates

1 overlap

Rearrangements (ABySS)

Pathway analysis and interpretation

Mutation frequency significance analysis (MutSig)
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

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Mutations</th>
<th>Patients</th>
<th>Unique sites</th>
<th>q-value no HM</th>
<th>q-value with HM</th>
<th>Expressed in 10 neuroblastoma transcriptomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALK</td>
<td>anaplastic lymphoma receptor tyrosine kinase</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>7.7x10^{-7}</td>
<td>2.6x10^{-6}</td>
<td>Yes</td>
</tr>
<tr>
<td>IKZF3</td>
<td>IKAROS family zinc finger 3 (Aiolas)</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>0.0047</td>
<td>0.0077</td>
<td>Yes</td>
</tr>
<tr>
<td>PSMD3</td>
<td>proteasome 26S subunit, non-ATPase, 3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0.0069</td>
<td>0.011</td>
<td>Yes</td>
</tr>
<tr>
<td>PGLYRP3</td>
<td>peptidoglycan recognition protein 3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0.045</td>
<td>0.065</td>
<td>No</td>
</tr>
<tr>
<td>LILRB1</td>
<td>leukocyte immunoglobulin-like receptor, subfamily B, member 1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0.071</td>
<td>0.085</td>
<td>Yes</td>
</tr>
<tr>
<td>PTPN11</td>
<td>protein tyrosine phosphatase, non-receptor type 11</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0.13</td>
<td>0.17</td>
<td>Yes</td>
</tr>
<tr>
<td>NRAS</td>
<td>neuroblastoma RAS viral (v-ras) oncogene homolog</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.17</td>
<td>0.17</td>
<td>Yes</td>
</tr>
<tr>
<td>GABRA6</td>
<td>gamma-aminobutyric acid (GABA) A receptor, alpha 6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1.00</td>
<td>0.17</td>
<td>No</td>
</tr>
<tr>
<td>SUCLG2</td>
<td>succinate-CoA ligase, GDP-forming, beta subunit</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1.00</td>
<td>0.17</td>
<td>Yes</td>
</tr>
<tr>
<td>IGSF11</td>
<td>immunoglobulin superfamily, member 11</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1.00</td>
<td>0.18</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Getz et al, Science, 2007
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)

- 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
  - \textit{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 \textit{ALK} mutants)
  - Mutations in chromatin remodeling genes in 1\% cases (some overlap with \textit{ALK} mutants)
  - 3\% of high-risk cases display a hypermutator phenotype that can be linked to mutations in DNA repair genes (e.g. \textit{MLH1})
- Coding sequence analysis suggests a potential therapeutic avenue in ~30\% of cases
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
-- Katayo Kasaian
-- 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 Insitutue
-- 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

The Jordan Hopkins Foundation
Whole Exome Sequencing at the Broad
In-solution hybrid-capture using RNA ‘baits’

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