# Integrative omics analyses for tumor stratification and CAR T cell therapy

Substantial differences in molecular profiles of pediatric and adult AML
 Identification of gliomas enriched for CAR T cell therapy

Hamid Bolouri

Division of Human Biology Fred Hutchinson Cancer Research Center TARGET

Therapeutically Applicable Research to Generate Effective Treatments

http://target.cancer.gov/

### NIH

Daniela Gerhardt Tanja Davidson,... FHCRC (pediatric AML) Soheil Meshinchi Rhonda Ries

Phoenix Children's (DNA.me)

Robert Arceci Jason Farrar, ... + contractors at St Jude's and BCCA

Bioinformatics Working Group: Tim Triche Jr, Jason Farrar, Emilia Lim, HB

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

### CHILDREN'S ONCOLOGY GROUP

The world's childhood cancer experts

Todd Alonzo Alan Gamis Rob Gerbing

### **Data Overview**

Pediatric Acute Myeloid Leukemia

- □ 197 whole genome sequences (matched Dx & remission)
- □ RNA-seq for 158 samples (500 more processed)
- □ miRNA-seq for 637 samples
- □ 142 Infinium 450K methylation arrays (~600 more to come)
- □ > 50 clinical data elements (incl. cytogenetics) per sample
- □ Targeted sequencing of ~400 genes in
  - 143 of the 197 whole-genome-sequenced samples
  - ~ 650 additional 'frequency validation' samples

*Plans/hopes* for 'full coverage' of a total of ~2000 samples.

junctions CNVs > 10Kbp Indels ~ 1 – 10 bp SNVs

### CompleteGenomics whole genome sequencing: 31- to 35-base mate-paired reads up to 700bp apart





### miRNA-seq (n=637)

### Emilia Lim



### Unsupervised clustering of DNA methylation probes (n=142)

(Tim Triche Jr and Jason Farrar)



Normal blood cells

variant verification

discovery set





Removed avgNormalizedCvg < 20 Removed standard deviation of LAF > 0.22 Removed ploidyScore < 30 Checked no chrM == OK

Removed CNVs in centromeric/telomeric regions

Removed 6 CNV regions < 10Kbp

Per patient, merged CNV regions within 10Kbp distance

Marked chr-arm level CNVs (footprint > 50%)

Added-in chr CNV data from CDEs

(115 of 197 samples have CNVs)

Matched the selected CGI CNVs (475 up, 488 down) to

### recurrent SNP6 CNV regions in 192 matched Dx samples:

- recur in at least 5 samples (2.5%)
- 803,917,882bp amplified in 894 segments
- 533,421,564bp deleted in 1434 segments
  - ➢ 402 (85%) match for CN up
  - > 367 (75%) match for CN down

15/15 predicted ELF1 deletions were confirmed by a nanoString tiling array





### Example sub-populations enriched for specific targets





Mutual exclusivity patterns



### Substantial differences in the genomic landscapes of pediatric and adult AML



Substantial differences in the genomic landscapes of pediatric and adult AML



### Independent confirmation



TARGET v TCGA

TARGET v 398 ECOG patients (Ross Levine, MSKCC)



Pediatric Acute Myeloid Leukemia (AML)

### **Failure** of a normal developmental process (block in HSC differentiation) + massive proliferation of immature white blood cells





Blood, 2005, (106):1519-1524

© 2011 Terese Winslow LLC



**HSC Differentiation Pathway** 



### Integrative omics analyses for

### tumor stratification and CAR T cell therapy

- 1) Stratification of AMLs by co-occurrence and mutual exclusion of events
- 2) Using multi-modal sample similarity to identify gliomas enriched for CAR T cells

(with Lue Ping Zhang and Eric Holland)

### 1105 TCGA gliomas

- Single Nucleotide Alterations (SNAs) from exome-sequencing
- Copy Number Alterations (CNAs) from SNP6.0 arrays
- DNA methylation from Infinium 450K arrays
- mRNA-seq
- Clinical data, but

 $\sim 2/3^{rd}$  of lower grade gliomas were 'alive' at data collection

 $\simeq 1/5^{\text{th}}\,$  have no status information





### feature sets ——

Approach:

- ✓ Use many similarity measures
- ✓ Use many feature (gene) sets
- ✓ Plot sample similarities in 2D
- ✓ Co-color & view sample subsets across all plot
- Use automated methods to find informative plots





## Example available gene sets (similarity/distance 'features')

- 🕨 H (hallmark gene sets, 50 gene sets) 🖬
- 🕨 C1 (positional gene sets, 326 gene sets) 🔽
  - by chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
- 🕨 C2 (curated gene sets, 4725 gene sets) 🔽
  - CGP (chemical and genetic perturbations, 3395 gene sets) 2
  - 🕨 CP (Canonical pathways, 1330 gene sets) 🖬
  - CP:BIOCARTA (BioCarta gene sets, 217 gene sets)
  - CP:KEGG (KEGG gene sets, 186 gene sets)
  - CP:REACTOME (Reactome gene sets, 674 gene sets)
- 🕨 C3 (motif gene sets, 836 gene sets) 🖬
  - MIR (microRNA targets, 221 gene sets) 2
  - TFT (transcription factor targets, 615 gene sets) 2
- 🖻 C4 (computational gene sets, 858 gene sets) 🖬
  - CGN (cancer gene neighborhoods, 427 gene sets) 2
  - CM (cancer modules, 431 gene sets) 2
- 🕨 C5 (GO gene sets, 1454 gene sets) 🖬
  - 🕨 BP (GO biological process, 825 gene sets) 🔽
  - 🕨 CC (GO cellular component, 233 gene sets) 🔽
  - MF (GO molecular function, 396 gene sets) 1
- C6 (oncogenic signatures, 189 gene sets) 2
- 🕨 C7 (immunologic signatures, 1910 gene sets) 🔽



### Example distance measures

#### Numerical Data

EuclideanDistance • SquaredEuclideanDistance • NormalizedSquaredEuclideanDistance

ManhattanDistance • ChessboardDistance • BrayCurtisDistance • CanberraDistance

CosineDistance 
CorrelationDistance BinaryDistance TimeWarpingDistance

#### **Boolean Data**

HammingDistance JaccardDissimilarity MatchingDissimilarity DiceDissimilarity RogersTanimotoDissimilarity RussellRaoDissimilarity SokalSneathDissimilarity YuleDissimilarity

#### String Data

EditDistance • DamerauLevenshteinDistance • HammingDistance • SmithWatermanSimilarity • NeedlemanWunschSimilarity

Images & Colors

ImageDistance 
ColorDistance

**Geospatial & Temporal Data** 

GeoDistance 

DateDifference

### http://www.wolfram.com

### Genomic sample similarity measures



Joint SNA:CNA similarity = **S**/sum(**S**) + **C**/sum(**C**)



Gliomas can be subdivided into 8 genomic subtypes

Bolouri, Zhao, Holland, PNAS 2016



Non-CIMP LGGs are GBM-like genomically and by survival.

Bolouri, Zhao, Holland, PNAS 2016





ROR1 is a plausible target in gliomas



Protein Cell 2014, 5(7):496–502

### A set of 329 genes segregates gliomas into high and low ROR1-expression groups











Cluster of 183 short-lived genomically highly similar tumors.

Expression data is available for 105 of these samples.

70 (67%) of 105 samples are selected ROR1-high samples.



### High ROR1 mRNA levels are tumor-specific and associated with glioma mesenchymal subtype







Our ROR1-based sample selection is highly concordant with the intersection of TCGA expression cluster 'LGr4' with the '*Mesenchymal*' and '*Classical*' expression subtypes.



### Our findings are supported by RNA-seq data from an independent UW GBM cohort



TCGA alone

TCGA + UW

81% of genomically-selected tumors (gold) are positive for 1 of 3 CAR T cells

