Connecting clinical and genetic data on the St. Jude Survivorship Portal

• Inter-departmental collaboration in 2019
  • Dr. Les Robison (Epidemiology and Cancer Control, SJLIFE)
  • Dr. Jinghui Zhang (Computational Biology)
• Built by the Comp Bio data visualization team led by Xin Zhou
  • 2019.10 ASHG → SJLIFE
  • 2020.7 AACR → SJLIFE and CCSS
  • Portal is under active development

• Design goals:
  • Curate and share clinical and genetic data from pediatric cancer survivors
  • Supports interactive real-time data analysis to promote genetic and epidemiology research
Clinical and genetic data

Clinical data

1531 variables about baseline assessments
- Cancer-related Variables
- Demographic Variables
- Self-reported Behavior and Outcome Variables
- Clinically-assessed Variables

7169 participants with clinical data
- CCSS n=2641, SJLIFE n=4528

Germline SNVs from whole-genome sequencing

90 million single nucleotide variations in total
5773 participants with WGS data
- CCSS n=2641, SJLIFE n=3132
The Survivorship Portal

About the Project
The Survivorship Portal shares high quality genomic, clinical, and patient-reported data from survivors of pediatric cancer. To accelerate the rate of discovery in survivorship research we have developed this SJLIFE Survivorship Portal, a data-sharing platform for genomic and clinical data from the St. Jude Lifetime Cohort hosted on the St. Jude Cloud. The Portal features the Clinical Data Browser, GenomePaint, a genetic variant browser, for browsing, visualizing and analyzing clinical and genetic data integratively. Additionally, the Survivorship Portal will serve as a site for an expanding portfolio of risk-prediction tools developed using the SJLIFE cohort, including the recently developed Cumulative Burden Risk-Prediction Tool.

Methylation beta value matrix for 513 CpGs. Reference: Epigenetic Age Acceleration and Chronic Health Conditions among Adult Survivors of Childhood Cancer. JNCI 2020

Beta Release
This initial version of the portal is read-only but future updates will allow authenticated users to create customized downloading of source data. For more information or help, please contact us.

Projects

Clinical Data Browser
The Clinical Data Browser enables interactive exploration of 1) Cancer-related variables such as diagnosis and treatment; 2) Phenotypic outcomes such as BPNC status; and 3) Phenotypic outcomes such as BPNC status.

Genomic Data Browser
The GenomePaint displays genome-wide SNP, indel, and copy number variants computed from WGS, allowing to display and filter variants found in a variety of functional categories.

Cumulative Health Burden
This web application visualizes estimated cumulative burden resulting from pediatric cancer and administered treatment. The calculation is represented based on the statistical model.
Browsing the clinical dictionary

- Genomic Profiling Status
- Cancer-related Variables
- Demographic Variables
- Self-reported Behavior and Outcome Variables
- Clinically-assessed Variables
Cross-tabulating two variables

- Central nervous system (CNS), n=596
- Acute lymphoblastic leukemia, n=303
  - Hodgkin lymphoma, n=260
  - Wilms tumor, n=212
- Non-Hodgkin lymphoma, n=199
  - Neuroblastoma, n=186
- Rhabdomyosarcoma, n=120
  - Osteosarcoma, n=111
- Ewing sarcoma family of tumors, n=96
  - Acute myeloid leukemia, n=72
  - Other leukemia, n=10
- MDS/Acute myeloid leukemia, n=6

Average dose to heart + TBI, cGy:
- not treated, n=1300
- unknown exposure, n=464
- exposed, dose unknown, n=6
- ≤3000, n=814
- >3000, n=57

# of patients

- Central nervous system (CNS), n=596
- Acute lymphoblastic leukemia, n=303
  - Hodgkin lymphoma, n=260
  - Wilms tumor, n=212
- Non-Hodgkin lymphoma, n=199
  - Neuroblastoma, n=186
- Rhabdomyosarcoma, n=120
  - Osteosarcoma, n=111
- Ewing sarcoma family of tumors, n=96
  - Acute myeloid leukemia, n=72
  - Other leukemia, n=10
- MDS/Acute myeloid leukemia, n=6

% of patients

- Central nervous system (CNS), n=596
- Acute lymphoblastic leukemia, n=303
  - Hodgkin lymphoma, n=260
  - Wilms tumor, n=212
- Non-Hodgkin lymphoma, n=199
  - Neuroblastoma, n=186
- Rhabdomyosarcoma, n=120
  - Osteosarcoma, n=111
- Ewing sarcoma family of tumors, n=96
  - Acute myeloid leukemia, n=72
  - Other leukemia, n=10
- MDS/Acute myeloid leukemia, n=6
Graded adverse events

- Self-reported Behavior and Outcome Variables
  - Graded adverse events
    - Auditory system
    - Cardiac system

Maximum grade of each patient

Cumulative incidence

<table>
<thead>
<tr>
<th>Grade</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: No condition</td>
<td>n=1114</td>
</tr>
<tr>
<td>1: Mild</td>
<td>n=105</td>
</tr>
<tr>
<td>2: Moderate</td>
<td>n=177</td>
</tr>
<tr>
<td>3: Severe</td>
<td>n=42</td>
</tr>
<tr>
<td>4: Life-threatening</td>
<td>n=42</td>
</tr>
</tbody>
</table>

CTCAE grade 3-5

Cumulative Incidence (%)

Time to Event (years)
Genetic data browser

Locus-specific association study

Germline SNV
288 variants

-log10 P-value

RefGene

ARID5B

Arms

ARID5B

n=376, view stats

GROUP 1
ALLELE FREQUENCY OF
CCSS
AND
DIAGNOSIS GROUP IS Acute lymphoblastic leukemia

GROUP 2
ALLELE FREQUENCY OF
gnomAD
Adjust race background

TEST METHOD
Fisher exact test

Restrict to European Ancestry
Planed work

Enhance Opportunities for Research by Providing Access to CCSS Phenotype, Genotype and Outcomes Data Through Development of a Cloud-based Data Analysis Ecosystem

- Includes adding phenotype data for all 25,665 participants and genetic data on an additional 5,900 participants.

- Access to both raw and summary data supported by a searchable ontology-based data dictionary
  - Longitudinal follow-up and detailed lab results

- Conduct cloud-based computing to analyze genomic and clinical data using sophisticated computational/statistical pipelines in real time
  - Polygenic risk score
  - Adjust for covariates in association/regression analysis

- Download customized data sets through controlled access

- Access to a portfolio of intuitive, field-tested visualization tools to explore data and apply complex risk prediction modelling in a secure cloud environment
Filtering

GENETICALLY DEFINE... IS European Ancestry AND (AVERAGE DOSE TO HEART + ... IS x ≥ 0 OR CUMULATIVE ANTHRACYCLINE... IS x ≥ 0)

Apply filtering

- Central nervous system (CNS), n=674
- Acute lymphoblastic leukemia, n=452
- Hodgkin lymphoma, n=311
- Wilms tumor, n=256
- Non-Hodgkin lymphoma, n=253
- Neuroblastoma, n=220
- Rhabdomyosarcoma, n=138
- Osteosarcoma, n=130
- Ewing sarcoma family of tumors, n=110
- Acute myeloid leukemia, n=77
- Other leukemia, n=14
- MDS/Acute myeloid leukemia, n=6

- Acute lymphoblastic leukemia, n=263
- Hodgkin lymphoma, n=255
- Central nervous system (CNS), n=242
- Non-Hodgkin lymphoma, n=172
- Wilms tumor, n=113
- Ewing sarcoma family of tumors, n=96
- Osteosarcoma, n=93
- Rhabdomyosarcoma, n=86
- Neuroblastoma, n=85
- Acute myeloid leukemia, n=59
- Other leukemia, n=12
- MDS/Acute myeloid leukemia, n=4
Planned work

- **Bring the complete CCSS cohort to St. Jude Survivorship portal**
- **Clinical data**
  - Longitudinal follow-up and detailed lab results
- **Genetic data**
  - WGS indel genotype calls, HLA typing, telomere length, haplotype
- **Analysis features**
  - Polygenic risk score
  - Adjust for covariates in association/regression analysis
  - Data download