Burkitt Lymphoma Genome Sequencing Project (BLGSP): Introduction

Daniela S. Gerhard[1], Bruno M. Grande[2], Nicholas B. Griner[1], Corey Casper[3], Sarah Gerdts[4], Abraham Omoding[4], Jackson Orem[4], Sam Mbulaiteye[1], Charles G. Mullighan[5], John T. Sandlund[5], Thomas Alexander[5], John Kim Choi[5], Jeremy S. Abramson[6], Thomas Gross[1] , Ariela Noy[7], Jeffrey Bethony[8], Fabio Leal[8], Timothy C. Greiner[9], Elaine S. Jaffe[1], Nancy L. Harris[6], Julie M. Gastier-Foster[10], Jay Bowen[10], Benjamin Hanf[10], Roland Schmitz[1], Jean Paul Martin[11], Marie-Reine Martin[11], John D. Irvin[11], Ellen Miller[12], Patee Gesuwan[1], Leandro Hermida[1], Tanja M. Davidsen[1], Karen Novik[13], Andy Mungall[13], Yussanne Ma[13], Marco Marra[13], Ryan Morin[13], Louis M. Staudt[1]

[1] National Cancer Institute, Bethesda, MD, [2] Simon Fraser University, British Columbia, Canada;[3] Fred Hutchinson Cancer Research Institute, Seattle, WA;[4] Uganda Cancer Institute, Kampala, Uganda;[5] St. Jude Children's Hospital, Memphis, TN;[6] Massachusetts General Hospital, Boston, MA;[7] Memorial Sloan Kettering Cancer Center, New York, NY;[8] George Washington University, Washington DC;[9] University of Nebraska Medical Center, Omaha, NE; [10] Nationwide Children's Hospital, Columbus, OH;[11] Foundation for Burkitt Lymphoma Research ;[12] Leidos Biomedical Research, Frederick, MD;[13] British Columbia Cancer Agency, British Columbia, Canada

Introduction

Burkitt Lymphoma (BL) is an aggressive B-cell lymphoma with a translocation involving *MYC* and immunoglobulin (Ig) loci. It is most common in children, but also affects adults, and occurs in sporadic, endemic and HIV-associated forms. The Epstein-Barr virus (EBV)-associated endemic subtype is the most common pediatric cancer in equatorial Africa yet also occurs in other parts of the world, for example in the Brazil rain forest. The endemic BL in Africa occurs at a 10-fold higher frequency than the sporadic subtype. Intensive chemotherapy is effective, but the associated toxicity requires supportive care that is not readily available in resource-poor regions. Previously published molecular characterization of small numbers of tumors indicated that the mutation profiles of endemic and sporadic cases are similar, but not identical.

Goals

1. To perform full molecular characterization (whole genome DNA sequencing, whole RNA transcriptome sequencing, miRNA sequencing) of 160 BLs (discovery set), 50% of which will be endemic, 38% sporadic (pediatric and adult) and 12% from HIV+ patients. Each case will get normal tissue DNA sequenced and will include clinical data such as treatment modality and outcome.
2. Analyze the molecular data generated to define the genetic and phenotypic features that drive these cancers with the intent toward developing new therapeutic strategies that can be deployed worldwide based on patients' molecular alternations.
3. Validate the molecular characterization results in an independent cohort of 100-200 cases (validation set)

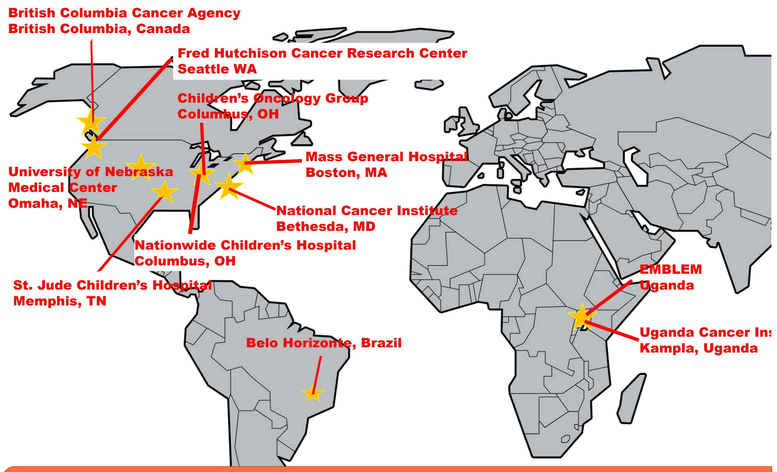
Project Requirements

Discovery cases meet the criteria outlined here:

- Tumor tissues from untreated patients diagnosed and confirmed by a group of three pathologists as BL
- 100 mg of flash-frozen (FF) tumor tissue is preferred
  - In lieu of frozen tissue, FFPEs may be used if the following requirements are met:
    - The block includes least 10-20 mg of tissue
    - Fixative buffer pH was in the neutral range
- Case-matched normal tissue, e.g. 10 mL blood, or 3 buccal swabs, or ~100 mg of normal tissue, or 5 ug of DNA
- Tumors have a minimum of 50% tumor nuclei and ~80% viable cells
- The patient did not receive neo-adjuvant therapy prior to tissue collection
- Treatment and outcome data is available
  - No prior diagnosis of a malignant neoplasm is preferred, or at least documented

Samples that do not meet these requirements will be used during the validation phase of the project to confirm and validate any detected genetic aberrations

BLGSP Collaborators



Sequencing and Data Analysis

- 80X and 40X coverage of the tumor and normal genome DNA, respectively
  - Allows for differentiation between rare germline variants and somatic mutations
- Transcriptome of tumor RNA and miRNA in depth high enough to verify somatic changes that are expressed at "medium" or higher levels
- DNA and RNA from ~30 BL cases isolated from FF and FFPE stored tissue will be sequenced in parallel to define correction parameters which would allow interpretation of FFPE-only derived sequence which are known to impact on technical aspects and therefore without the empirical data could lead to artefacts
- Raw sequencing data (BAM, FASTQ files) will be deposited into the Genomic Data Commons (GDC, gdc.cancer.gov), designed to integrate molecular and clinical data for NCI large scale cancer genomic projects in a single data service
- Analyzed data (both clinical and molecular) will be available through the Data Coordinating Center (DCC) (http://cgci.nci.nih.gov/dataMatrix/) with both open and controlled access to the protected tier (identifiable clinical and genetic data) to users that agree to the data access policy
  - Translocations were identified with 2 methods, MANTA and DELLY and the results are compared; chromosome copy number was detected by Sequenza
  - Somatic point mutations and insertion/deletions are detected by STRELKA

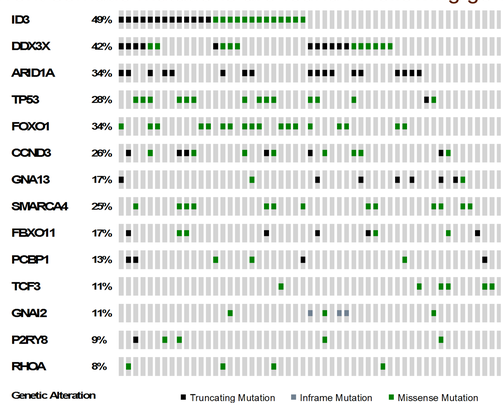
For more information, please visit the BLGSP website: https://ocg.cancer.gov/programs/cgci/projects/burkitt-lymphoma

Results Summary

- **Tissue Procurement Status**
  - To date we identified eight Tissue Source Sites (TSSs), two in Africa, one in Brazil and five in the USA
  - To aid the need for technical reproducibility, but still be able to work with the various TSSs, we developed Standard Operating Procedures, https://ocg.cancer.gov/resources/all-resources/#161
  - 157 cases were identified by five TSSs; some continue to accrue, while three TSSs are being brought into the project
    - 69 passed all criteria of discovery
    - 29 passed criteria for validation
    - 30 are awaiting pathology review or nucleic acid extraction
    - 29 failed either pathology or molecular quality criteria
- **Lessons Learned**
  - Strict adherence to SOPs are required, otherwise artificial "batch effects" are detected upon analysis of the molecular data
- **Preliminary Results from Analysis of Sequence Data**

Note: The experiments are ongoing so the # of cases per analysis varies

- The BL genome is relatively quiescent (N=37)
  - Gain of 1q21 is present in ~20% of the cases
  - Interstitial deletion in 11q is found in ~5% of the cases
- 98% of cases have the MYC/Ig translocation(s); the one case without has 11q interstitial deletion previously found in a few rare BL cases and another case has both t(MYC/Ig) and 11q deletion
- 77% of 53 cases express Epstein Barr Virus transcripts
  - From Africa (N=43) 91% are EBV^{+} and 9% EBV^{-}
- Recurrent somatic mutations are found in following genes (N=53):



- We found rare germline variants in CCNF, but no somatic mutations

- Integration will be performed when the DNA and RNA sequencing of the discovery cases is complete