# FOUNDATION FOR BURKITT LYMPHOMA RESEARCH

BURKITT LYMPHOMA GENOME SEQUENCING PROJECT (BLGSP) 11/24/15

## EXECUTIVE SUMMARY

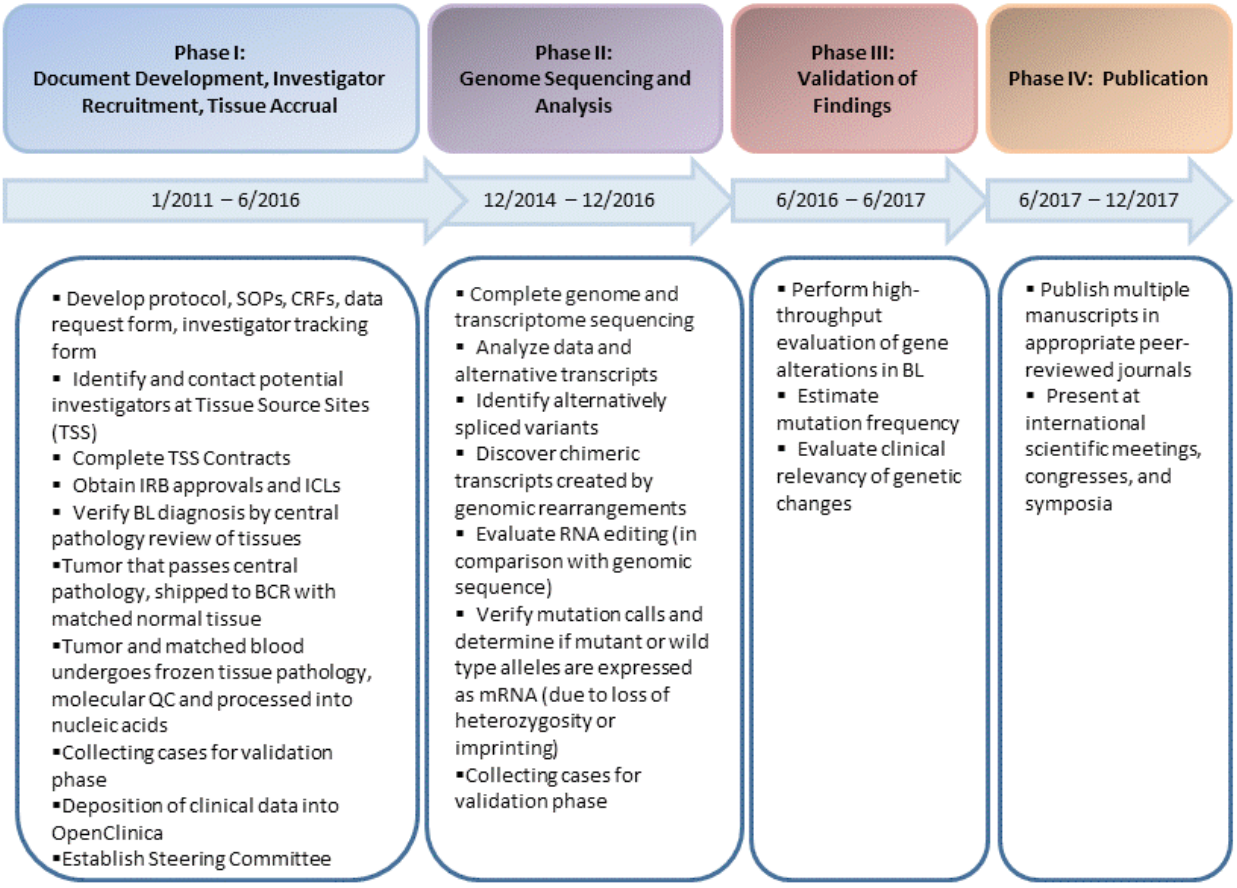
The Foundation for Burkitt Lymphoma Research (FFBLR) has established a collaborative effort with the National Cancer Institute (NCI), in order to develop a genomic databank for Burkitt Lymphoma (BL). The BLGSP will compile genetic changes present in BL tumors, analyze the data to identify diagnostic, prognostic, or therapeutic markers or targets, and publish the results. As with other cancer sequencing projects, the goal is to identify potential genetic changes in patients with BL that could lead to better prevention, detection and treatment of the cancer.

The effort will be coordinated through the Foundation for the National Institutes of Health (FNIH). The NCI, through the Office of Cancer Genomics (OCG), will provide the administrative, physical and analytical infrastructure in order to carry out this project. This will enable the FFBLR to utilize existing Standard Operating Procedures (SOPs) developed by OCG, as well as relationships OCG has established with other organizations, in order to carry out this project. This effort will enhance the scientific integrity and credibility of the BLGSP, and be consistent with the conduct and methodologies used in other NCI sponsored cancer sequencing projects.

The FFBLR will assist in the accrual of the tissues by identifying sites in the US, Europe and Africa, making the initial contacts and providing administrative support. NCI will establish the contracts and protocols for accrual and quality control that are already in place as part of ongoing projects, such as the Cancer Genomic Characterization Initiative (http://ocg.cancer.gov/programs/cgci). The FFBLR will participate in the scientific leadership of the BLGSP, including having members of its Scientific Advisory Board review the data from the project; provide the biologic context for publication and dissemination within the appropriate scientific communities.

It is anticipated that the project timeline will be 3-4 years: approximately 2 years to identify and collect appropriate tissue and data from approximately 160 qualifying patients with BL (Phase I), about 6-12 months to complete the sequencing and analysis (Phase II), 6 months for validation studies (Phase III), and another 3-6 months to publish the data (Phase IV).

# Burkitt Lymphoma Genome Sequencing Project Overview (updated 11.24.15)



# INTRODUCTION

Dr. Jean Paul Martin and Dr. Marie Reine Martin have established a Foundation for Burkitt Lymphoma (BL) Research in Geneva, Switzerland to further the understanding and the treatment of BL. They made this decision because their son Xavier was diagnosed in July 2009 with BL and passed away from a relapse of the disease in May 2010.

Burkitt Lymphoma, first described by Dr. Denis Parsons Burkitt in 1956, is an uncommon type of Non-Hodgkin Lymphoma, often but not exclusively affecting children. It is a highly aggressive type of B-cell lymphoma and often involves body parts other than lymph nodes. The disease is associated with a chromosomal translocation of the *MYC* gene. The tumor consists of sheets of a monotonous (similar in size and morphology) population of medium size lymphoid cells with high proliferative activity and apoptotic activity described as a "starry sky" appearance under low power magnification.

Currently BL is divided into three main clinical variants: 1) Endemic: This BL variant occurs mainly in equatorial Africa, where the disease was first described and is the most common malignancy in children. In 95% of cases, the children are infected with Epstein-Barr virus, which is presumably key to its etiology. The disease involves children much more than adults. The disease characteristically involves the jaw or other facial bone, and may involve the distal ileum, cecum, ovaries, kidney or the breast. 2) Sporadic: This BL variant is found outside of Africa and is usually not associated with Epstein-Barr virus infection, but has morphological and molecular features in common with endemic BL. The jaw is less commonly involved, whereas the ileocecal region is the most common site of involvement. 3) Immunodeficiency-associated: This variant is usually associated with HIV infection and can be associated with the initial manifestations of AIDS, but can also occur in post-transplant patients who are taking immunosuppressive drugs.

Current chemotherapy regimens are effective in approximately 40-90% of patients depending on age, stage of the disease, treatment regimen, and site of the treatment facility. Hence, new treatments are needed to improve the efficacy of current regiments and potentially to substitute less toxic agents for the high intensity chemotherapeutic drugs that are currently given.

With this as a brief background of BL, the vision of the Foundation is to provide monetary support of prioritized, key pre-clinical and clinical research activities that will provide a better understanding of the mechanisms of the disease and lead to more effective treatment algorithms.

The Foundation has a renowned international Advisory Board including Professor Jane Apperley (Emeritus Member), Dr. Corey Casper, Professor Riccardo Dalla-Favera, Professor Gerard Evan, Dr. Charles Mullighan, Dr. Ariela Noy, and Dr. Louis Staudt. The Executive Director of the Foundation is Dr. John D. Irvin.

The Foundation's first key research project is to establish a comprehensive databank of genomic and clinical data in patients with BL. The purpose of this document is to outline and describe the activities associated with developing the database.

# AIM

This project will establish a comprehensive databank of genomic sequence from patients with BL, utilizing tumor DNA and matched normal DNA as well as tumor RNA. Clinical data will also be collected so that associations between clinical parameters and genetic abnormalities can be discovered.

## SIGNIFICANCE

Since the discovery in 1983 of *MYC* translocations in BL, little progress has been made in understanding the molecular pathogenesis of the disease. MYC, the defining oncogene for this cancer, is itself insufficient to transform primary cells and actually induces apoptosis unless growth factors are provided or cooperating oncogenes are activated. Loss of p53 function by mutation or by inactivation of INK4a$^{ARF}$ may contribute to cell survival in a fraction of cases. However, there has yet to be a comprehensive effort to discover new genetic aberrations associated with BL using high throughput methods.

There are several major unmet needs in the clinical management of BL that require urgent attention. First, current chemotherapy is ineffective in roughly 20% of cases, and is associated with therapy induced toxicity and death. Because of the toxicity, patients with BL over the age of 70 are ineligible for this potentially curative therapy. Second, young individuals exposed to these intensive regimens face the real possibility of secondary malignancies years later. Finally, children with endemic BL in Africa in general do not receive these curative chemotherapy regimens because of the lack of hospital support for the management of post-treatment infections, and consequently most will die of their disease.

What is needed, therefore, are new approaches to the therapy of BL that are based on an understanding of its molecular pathogenesis and the regulatory pathways that it utilizes for proliferation and survival. Currently, we have no insight into the possible involvement of kinases and signaling pathways that might promote the high rate of proliferation that characterizes these tumors. Likewise, mechanisms that prevent apoptosis of BL cells, besides p53 inactivation, have yet to be discovered.

Burkitt Lymphoma is derived from the germinal center stage of B cell differentiation, and as such expresses a professional mutator, activation-induced cytosine deaminase (AID). AID is likely involved in generating the breakpoints that lead to MYC translocations in sporadic BL and almost certainly targets many cellular genes for mutation. Hence, BL may utilize AID-generated mutations to modulate oncogene and tumor suppressor pathways. This BL comprehensive genome characterization project will shed considerable light on the oncogenic capacity of AID and identify other potential disease-causing candidates.

An additional fascinating question is whether genetic alterations in the BL genome may influence the symbiotic relationship with the Epstein-Barr virus in the endemic variant in a proportion of the sporadic and immunodeficiency-related cases. EBV viral proteins can mimic Notch signaling (EBNA2), CD40 signaling (LMP1), and B cell receptor signaling (LMP2). It may be necessary for the BLs to acquire mutations in these pathways to shape the cellular response to these virally derived signals.

Many key clinical questions may be illuminated by knowledge of genomic abnormalities in BL. Why do approximately 10-60% of patients treated with potentially curative chemotherapy nonetheless succumb to their disease? How do the genomic alterations in the BL variants (endemic, sporadic, immunodeficiency-related) differ, and how do these differences affect their response to therapy? What new therapeutic options may be available to augment or supplant high-dose chemotherapy?

## STUDY POPULATION

The study population should include fresh tissue samples (case-matched tumor and normal) with the possibility of FFPEs from 30 adult patients with sporadic BL, and tissue from approximately 130 patients of each of the other three variants, pediatric sporadic, endemic, and HIV-associated BL. It is anticipated that about double that number of cases will need to be collected as there is about a 50% failure rate at the pathology and molecular processing stages.

## STUDY CONDUCT AND STANDARD OPERATING PROCEDURES

In order to enhance the scientific integrity and credibility of the Burkitt Lymphoma Genome Sequencing Project (BLGSP), the project will be carried out utilizing SOPs developed by the OCG at the NCI. The BLGSP-specific SOPs can be found in the OCG Tumor Molecular Characterization Projects Standard Operating Procedures manual (https://ocg.cancer.gov/sites/default/files/BLGSP_SOP_manual_09252015_v3.pdf) and are consistent with those currently utilized in the NCI-sponsored HIV+ Tumor Molecular Characterization Project (HTMCP).

Using the SOPs, the tissues will be accrued by collaborators (referred to as Tissue Source Sites [TSS]) with access to BL patients. Each TSS will need to obtain an institutional review board (IRB)-approved study protocol that allows for collection of human tumor tissue and matched normal tissue (preferably blood). Upon approval, all TSS will be required to collect and store the tissues, perform site-capable pathology analysis to determine the BL diagnosis, and provide the BLGSP-required clinical data. The above-referenced Burkitt Lymphoma Genome Sequencing Projects SOP Manual includes templates of an informed consent and an institutional certification letter, which can be adapted by each TSS for their use. Every patient studied in the BLGSP must be enrolled in a site-specific protocol, and agree to participate by signing an informed consent. The institutional certification letter states that the IRB understands that the data generated will be made publicly available with the appropriate patient protection procedures (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/about.html).

## TISSUE COLLECTION AND PREPARATION

Procedures for tissue collection, preparation, pathology review and shipment in the BLGSP are detailed in the Burkitt Lymphoma Sequencing Project SOP Manual. Adherence to the SOPs is essential to ensure that the quality of the tissues collected meet the technical requirements of the methods of analysis. The protocols include information on tissue collection and preservation by freezing, pathology review, collection of blood samples for molecular characterization, shipping cryoports containing frozen biosamples for processing and extraction of nucleic acids, sample shipping guidelines, sample identifier standards, and disposition form for remaining macromolecules/tissues that are contributed to the project. Templates for suggested language for prospective tissue collection and the data release policy are also included within the SOPs.

## TISSUE REQUIREMENTS

Following is a summary of the tissue requirements for the BLGSP. The specific detailed requirements are included in the above referenced SOPs.

- Paired tumor and matched normal (non-involved tissue or blood) of untreated cases must be available in sufficient quantities.
- If blood is the normal tissue, documentation of normal peripheral white blood cell counts and smears must be provided since in rare instances, BL cells can be in the peripheral blood. There are 2 ways of getting normal tissue without tumors: a) Granulocytes can be separated from lymphocytes with Ficoll. b) Also with Ficoll separation, flow cytometry analysis can be performed with CD10, a BL surface marker.
- In lieu of frozen tissue, FFPEs may be used if it meets the following requirements:
    - At least 10-20 mg of tissue

    - Fixative buffer pH should be noted
    - No age requirement, but age of FFPEs should be noted
- Pre-treatment blood sample is preferred as the source of normal tissue. However, if a tumor specimen is pre-existing and the patient is alive and can be re-consented following treatment for the collection of a post-treatment blood sample, then that will be acceptable.
- Tissues (both normal and tumor) need to be snap frozen, if frozen tissues are to be submitted. Currently the project anticipates that ~100 mg of tumor tissue and 10 ml of blood is required, but this may change as the sequencing characterization technology improves.
- A portion of tumor adjacent to the snap frozen sample will be assessed to estimate % tumor cells, with additional immunohistochemical (IHC) stains and FISH to be performed to validate the diagnosis, as needed (see below). Tumors need to have minimum 50% tumor nuclei as assessed on the H&E section from the frozen tissue with ~80% viable cells.
- Standard World Health Organization diagnostic criteria for BL must be fulfilled. Specifically, the tumor cells have the "starry-sky" appearance, be positive for CD10 and/or BCL6 by IHC and >90% of the cells need be Ki67 positive. It would be optimal if the presence of MYC to immunoglobulin locus translocation was confirmed by fluorescent in situ hybridization (FISH). At least 12 paraffin-embedded slides, or even better a block, have to be provided, from each case for pathology review by the project pathologists.
- Clinical data (see SOPs) has to be collected at the time of diagnosis, updated after treatment, and then every year for the next 2 years.

If tumor tissue or blood/constitutional tissue is not available, but clinical information and paraffin-embedded slides or block(s) are, we would consider accepting nucleic acids from pathology review-confirmed cases that meet the following criteria:

- At least 10 µg of RNA with RNA Integrity Number (RIN) > 7 as determined on an Agilent Bioanalyzer or similar system
- At least 5 µg each of tumor and normal high molecular weight (>80% of the molecules are >10 kb) DNA

The tissues will be processed in a single biospecimen-processing center using standard protocols developed for and validated in Therapeutically Applicable Research to Generate Effective Treatments (TARGET)( http://ocg.cancer.gov/programs/target) and The Cancer Genome Atlas (TCGA)( http://cancergenome.nih.gov/) The RNA and DNA will be quality controlled and the match of the tumor and normal DNA confirmed. As part of the processing, the DNA will be whole genome amplified by the method considered the most optimal for genomic characterization at the time when this project will be ready to do so. One of the BLGSP's outcomes will be the availability of the DNA and/or RNA for other projects.

## PATHOLOGY REVIEW/VERIFICATION

The review of all tissues by a central pathology review panel of three board-certified hematology pathologists will ensure that the samples that are molecularly characterized are BL, and remove the subjectivity encountered in standard pathology practice. The pathology criteria for BL diagnosis include: 1) characteristic histological appearance; 2) Ki67 immunohistochemical staining in >90% of tumor nuclei; 3) CD10+ and/or BCL6+ by immunohistochemistry; 4) *MYC/Ig* translocation by FISH analysis; BCL2 staining by IHC and those cases which are positive, will by analyzed by *MYC/BCL2* FISH. A SOP for this review and verification has been established and is available in the BLGSP section of the OCG SOP manual.

## SEQUENCING, ANALYSIS AND VALIDATION

One of the major objectives of the BLGSP will be to generate genomic-scale DNA and/or RNA sequence from case-matched tumor-normal pairs from patients with BL. For DNA as well as RNA sequencing, normal DNA needs to be sequenced to define which novel sequence variants are germline and which are somatic changes. The plan is to initiate sequencing once a critical number of cases (12-20) have been collected and continue sequencing in batches throughout the term of this project. We anticipate that the cost of whole genome sequence will be decreasing, though the exact cost cannot be estimated. The analysis of the genomic DNA sequence will provide information of single-nucleotide polymorphisms (SNPs) for loss of heterozygosity, somatic mutations, genomic rearrangements and copy number.

We anticipate that sufficient funds will be available for whole transcriptome sequencing (mRNA-seq and miRNA-seq). The analysis of that data results in digital gene expression profile, identification of alternative spliced variants, discovery of fusion transcripts created by genomic rearrangements, RNA editing (in comparison with the genomic sequence), confirmation of mutation calls, and if mutant or wild type alleles are expressed as mRNA (either due to loss of heterozygosity or imprinting).

While the cost and quality of sequencing ($2^{nd}$, $3^{rd}$ generation) in two years is impossible to predict (currently the costs are decreasing rapidly and the quality of the sequence is improving as well), there is every reason to anticipate that it will be the norm to sequence the genome more than the current 30X (50-100X?) and include the transcriptome as well.

The current proposal does not include epigenetic characterization, however whatever materials will be left over could be used in the future to do genome-wide characterization of the methylation profile.

The NCI project team will develop an analytical protocol for the data generated, including:

- Analysis of the copy number alternations with integration with digital gene expression
- Identification of somatic mutations
- Identification of translocations

Taken together, these analyses will look for recurrent lesions within BL as a whole and within the BL variants. Genetic changes will be placed into biological pathways and integrated using available methods.

The findings will be validated in a new population cohort via high-throughput evaluation of gene alterations in BL as a whole and within the BL variants. The validation cohort will include up to 200 patients; the proportion of subtype cases will be selected based on the results obtained in the discovery phase (II). Validation studies performed in this new cohort will allow for mutations uncovered in Phase II to be verified as true and increase the power to estimate the frequency with which they occur. Moreover, validation will enable evaluation of the clinical relevancy of the genetic changes found.

## GOVERNANCE

The tissue contributors will be co-authors on the first manuscript and may collaborate in the future in others.

The project will have a steering committee (SC) whose membership includes members of the FFBLR's scientific advisory board (Dr. Louis Staudt; Professor Riccardo Dalla-Favera; Dr. Ariela Noy; and Dr. John Irvin, Executive Director FFBLR), one representative each from the contractor(s) doing the sequencing and tissue processing, the members of the pathology review board and 1-2 representatives from OCG as

needed. During the first 2 years of the project, the SC will meet by teleconference quarterly, and face-to-face annually to assess the progress in tissue accrual and any issues, which may arise in the process to isolation of the molecular analytes. The sequencing will be performed by contractors, which will be put in place by OCG. Once sequencing starts, the SC needs to meet monthly through teleconferencing and face to face every 6 months.

## DATA RELEASE POLICY AND PUBLICATION

The data that are generated will be submitted to publicly accessible databases developed for HTMCP and other projects. The NCI database is called the Data Coordinating Center and it allows an easy access to all data generated (TARGET, CGCI and TCGA) as well as analytical tools developed for other large-scale sequencing projects. The BLGSP will be listed on the CGCI web page and the Foundation acknowledged as a collaborator. The procedures developed for the other projects include submission of the sequence generated to publicly accessible databases (currently called sequence-read archive), which are an NCI-defined resource used by all NCI projects. The BLGSP agrees to abide by the data release policy formulated for the CGCI (http://ocg.cancer.gov/programs/cgci/using-cgci-data). Every effort will be made to rapidly, and not later than in 3-6 months after validation, publish the first manuscript describing the data and including a high-level analysis.