**A data analysis and visualization processing tool for Quantitative Multiplex Co-Immunoprecipitation(QMI) Platform**

**Feiping Li**

Master of Science in Computer Science & Software Engineering

**University of Washington  
6/8/2018**

**Project Committee:**Professor Wooyoung Kim, Committee Chair  
Professor Yang Peng, Committee Member  
Professor Geetha Thamilarasu, Committee Member

**Table of Contents**

[1 INTRODUCTION 3](#_30j0zll)

[1.1 Protein-Protein Interactions and Cancer Treatment 4](#_1fob9te)

[1.2 The Quantitative Multiplex co-immunoprecipitation platform (QMI) 5](#_3znysh7)

[1.3 Outline 7](#_2et92p0)

[2 RELATED WORKS 7](#_tyjcwt)

[2.1 Protein-Protein Interactions, T-Cells and disease mechanisms 7](#_3dy6vkm)

[2.2 Immune system, T-cells and immunotherapy 9](#_1t3h5sf)

[2.3 Quantitative Multiplex Co-Immunoprecipitation Platform facilitates T-cells research 12](#_4d34og8)

[2.4 A data analysis and virtualization tool for QMI platform 14](#_2s8eyo1)

[3 METHODOLOGY/SYSTEM DESIGN 15](#_17dp8vu)

[3.1 Requirements 15](#_3rdcrjn)

[3.2 System Architecture 15](#_26in1rg)

[3.3 Design Patterns 17](#_35nkun2)

[3.4 Technologies and Tools 18](#_1ksv4uv)

[3.5 Software Development Lifecycle Process 20](#_44sinio)

[3.6 Analyses 20](#_2jxsxqh)

[3.7 Measures 21](#_z337ya)

[4 RESULTS 22](#_3j2qqm3)

[4.1 Running an QMI analysis 22](#_1y810tw)

[4.2 Evaluation 25](#_4i7ojhp)

[5 CONCLUSION & FUTURE WORKS 26](#_2xcytpi)

**Abstract**

The project aims to provide an intuitive data analysis and interactive visualization processing program that facilitates the Quantitative Multiplex co-immunoprecipitation (QMI) platform [1]. The QMI platform is a novel approach that generates protein-protein interactions to provide medical practitioners with necessary data for engineering T-Cells, that can help patients’ immune systems fight against abnormal cells. However, the existing QMI platform has poor usability that lacks visualization and interactive interfaces, which necessitates an integration of scattered tools into a unified graphical user interface program. Here, we build a graphical user interface (GUI) based QMI analysis software program, which provides a streamlined workflow and an intuitive and interactive interface by integrating three separate analytic programs into a single program. We believe that the final product will have a huge impact on medical researches that require complex data analysis processes.

# INTRODUCTION

*Protein-protein interactions(PPIs) [4]* are of outstanding interest in biomedicine. Protein interaction networks form the molecular logic circuits for cells to make a cellular decision to switch from one homeostatic program to another. By interpreting the system-level behavior of PPI networks, researchers can understand the molecular circuitry underlying cellular behavior and the mechanisms of existing drugs. This enables researchers to chart the development and progression of disease states and to design effective drugs.

*Quantitative Multiplex co-immunoprecipitation (QMI)* [1] is a novel approach that facilitates protein-protein interactions to design customized medical plans for patients. However, due to its lack of usability and visualization, QMI is difficult for end users, such as lab technicians, to leverage and perform the step-by-step analyses. To facilitate the integration and validation of the QMI platform into the end user labs, our project aims to build a data analysis and virtualization software program for the QMI platform, providing an intuitive graphical user interface to simplify operations and present meaningful and visualized data and results.

## Protein-Protein Interactions and Cancer Treatment

PPIs sheds light on disease mechanisms. A protein’s function can be defined by the interactions with other proteins or the complexes in which it participates as a member. Indeed, a central paradigm of cell signaling views proteins and their interactions as a framework that mediates the cell’s capability to sense and respond to changes in its environment [12]. Studies [16] [17] [19] [20] have demonstrated that disordered protein networks are the main reason causing a genetic disease.

By studying protein-protein interactions, researchers try to find effective treatments for genetic diseases including cancer. Since the human immune system does not identify cancer cells, it fails to attack or kill them. For years, [chemotherapy](http://youdao.com/w/chemotherapy/) [13] has been a standard treatment for cancer. C[hemotherapy](http://youdao.com/w/chemotherapy/) usually targets fast-growing cancer cells, but it also kills fast-growing normal cells with the consequences that patients suffer from debilitating side effects. Recently, an alternative treatment, immunotherapy [14], has emerged and been validated by researchers [21] [22] [23]. Immunotherapy strengthens the power of a patient’s immune system to identify and kill cancer cells by designing and activating engineering T-cell receptors in a patient’s body.

T-cells, which is one type of immune system cell, are critical mediators that identify foreign substances. The immune system recognizes foreign substances by checking the contents of antigens on the surface of cells. T-cell receptors attach to foreign antigens and trigger an immune response to destroy the foreign cells. Figure 1 shows how T-cells examine a cell in human’s body. Unfortunately, because T-cells lack a specific type of T-cell receptor, the immune system often fails to identify cancer cells’ antigens. By activating the engineering T-cells in the body, researchers can build an adaptive immune system that is elegantly choreographed to recognize and attack cancer cells, just as it attacks viruses and other pathogens.

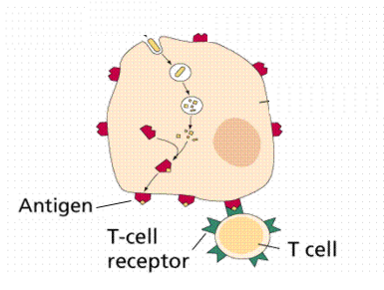


Figure 1: A T-cell uses T-cell receptors to examine antigens presented on the surface of a cell. [24]

To customize effective engineering T-cells for each patient, researchers collect samples from the patient and test them with different proteins, in order to find the combination that produces the most effective T-cells. By discovering the best combination of proteins, researchers can develop individually tailored drugs. Immunotherapy opens a door to advanced treatments that keep normal cells intact and cause minimal side effects.

Therefore, investigating of the dynamic molecular processes that involved in T-cell receptors (TCR) activation will improve researchers’ understanding of cancer pathogenesis and enhance their ability to control T-cells activation in the context of modern, cutting-edge immunotherapy treatments such as immune checkpoint inhibition [16] and chimeric antigen receptor T-cells [15]. Chapter 2 explains the two treatments in detail.

## The Quantitative Multiplex co-immunoprecipitation platform (QMI)

*Quantitative Multiplex co-immunoprecipitation (QMI)* [1], proposed in 2015 by Dr. Stephen E. P. Smith, an assistant professor of pediatrics at the University of Washington. QMI successfully performs hundreds of parallel co-immunoprecipitations [13] from small amounts of input biomaterial, generating “network biosignatures” that reflect the activation state of the T-cells (see figure 2). The SEPS lab [11], led by Smith, aims to integrate and validate the QMI platform into two clinical cancer research groups at two sites, the Mayo Clinic in Minnesota and Seattle Children’s Hospital in Washington states. The adoption of the QMI system, however, has faced several hurdles.

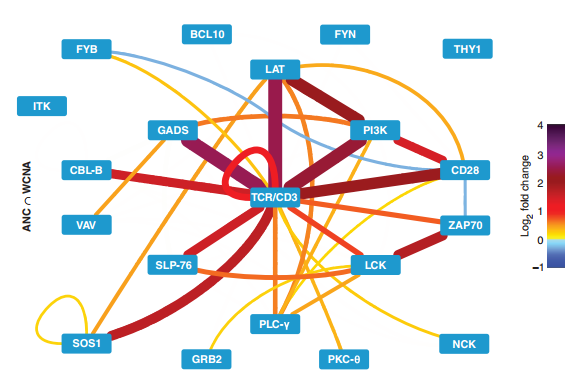


Figure 2: “network biosignatures” that generated from the QMI analysis. It reflects the activation state of the T-cells [1].

First, the QMI analytic process is complicated for end users. Even though the QMI system is described in detail in terms of protocols, identifiers, and mathematical equations for statistical analysis, QMI is still difficult to understand.

Second, no streamlined process and a simple interface is available. The existing QMI platform is implemented with three separate programs (Excel, MATLAB, and R). The workflow requires a user to export data generated by a Bioplex 200 instrument, convert them to different formats, and manually input and manipulate data in command-line interfaces in R and MATLAB, which necessitates programmers’ knowledge and input.

Third, the QMI platform lacks virtualization to make data and results meaningful and understandable for end users. Since it presents results in an unorganized manner, lab technicians would feel difficult to analyze the data.

To overcome these challenges, our project builds a data analysis and virtualization tool, which provides a streamlined bioinformatics workflow and an intuitive graphical user interface (GUI). It will integrate three statistical analysis programs into a single easy-to-use program which will facilitate the integration the QMI platform into end users’ labs.

Our project has three advantages. First, end users will be able to run a QMI analysis with little or no knowledge of statistical analysis, programming, and protocols. By having a graphical user interface running on top of different programs, end users can easily upload experiment data and perform all statistical analyses by clicking through it. Second, end users will find data and results meaningful and understandable though the virtualization provided by our project. Third, with consideration of usability, availability, performance, security and maintainability, our project will provide a satisfying user experience.

## Outline

The paper is organized as follows. Chapter 2 covers background about the protein-protein interactions (PPIs), T-cells research, and the QMI platform. Chapter 3 discusses the design, approach, technologies, and tools used for our project. Chapter 4 presents the results and demo of our project. A conclusion and the future phase of our project are presented in Chapter 5.

# RELATED WORKS

We first describe why protein-protein research is important for cancer research and treatment, then, we explain the function of T-cells and why the key role they play in immunotherapy. Next, we introduce the Quantitative Multiplex Co-Immunoprecipitation (QMI) platform [1] and its contribution to immunotherapy research. Finally, we discuss how our project can facilitate in the validation and integration of the QMI platform into end user labs.

## Protein-Protein Interactions, T-Cells and disease mechanisms

Protein-protein interactions (PPIs) are the physical and selective contacts that happen between a pair of proteins, in certain molecular regions and in a defined biological context. Human body is made out of cells and each living cell is packed with proteins that continuously interact with each other to control the cell's growth and eventual fate.

Because proteins’ functions are dependent on each other, proteins rarely act alone. They continuously cross talk with each other by the signaling processes. They resemble an organized army, by interacting with each other, spreading their thoughts, and following given commands. [4]

PPIs illuminate disease mechanisms. By comparing normal samples and diseased tissue samples, researchers have revealed that protein networks are perturbed in disease due to sequence mutations and expression changes. For example, Zhong [16] focused on known mutations causing Mendelian disorders. His study showed that nonsense mutations cause the node-removal effects in the PPI network, while in-frame mutations are associated with edge-specific perturbations. Subsequent studies by Wang [17], Wei and Yu [18], and Rolland [19] pointed out that disease mutations were appeared in perturbed PPIs that are functionally relevant in the particular tissue affected by the specific disease. Also, some studies show that disordered protein networks are the main reason that caused the genetic diseases [20].

Cancer is a genetic disease that is also caused by perturbed interactions. While cancer has many causes, ultimately all these causes exert their effects on a special class of genes called cancer genes or proto-oncogenes [26].  According to Griffiths [26], Oncogenes normally carry out basic cellular functions, generally related to the regulation of [cell division](https://www.ncbi.nlm.nih.gov/books/n/iga/A4529/def-item/A4637/). However, several types of events can change a [proto-oncogene](https://www.ncbi.nlm.nih.gov/books/n/iga/A4529/def-item/A5262/) into an oncogene—that is, into a state in which it promotes the two main characteristics of cancer: (1) uncontrolled cell division leading to an overgrown group of cells called a tumor and (2) the spread of tumor cells throughout the body to form new tumors, called metastasis. One of the ways in which proto-oncogenes can be changed into their cancer-causing (oncogenic) state is by mutation. Spontaneous or environmentally induced mutation occurs in a proto-oncogene of a single cell, which then undergoes multiple cell divisions to form a tumor. Because all the cells of the tumor carry the mutated oncogene, we can see that a tumor is a [mutant](https://www.ncbi.nlm.nih.gov/books/n/iga/A4529/def-item/A5113/) [clone](https://www.ncbi.nlm.nih.gov/books/n/iga/A4529/def-item/A4673/). [21]

For years, the cancer treatment includes Chemotherapy [27], along with surgery and radiation therapy. However, most of times, side effects of chemotherapy, which prevent fast-growing cells from dividing and growing, make patients suffer more. Moreover, Chemotherapy damages the healthy white blood cells that are normally the backbone of human’s immune system and are constantly replenished. A reduced white blood cell count, which is a common side effect of chemotherapy, damage or even shut down patients’ immune systems, leaves patients susceptible to infections, including those caused by bacteria, viruses, or fungi. [27]

## Immune system, T-cells and immunotherapy

Due to this severe side effects of Chemotherapy, researchers have strived to find advanced treatment for cancer. Immunotherapy [14], developed from studies on the PPIs and T-cells, aims to build an adaptive immune system that is elegantly choreographed to recognize and attack cancer cells, viruses, and other pathogens that are foreign to the body.

Immune system and T-cells works together to detects and destroys virus, fudges and other foreign cells that are harm to human’s body. T-cell alerts the immune system to mount an attack though a cascading series of responses. When a virus infects the host cell with its genome, by commandeering hidden inside, cells in human body use an intricate mechanism to extract some of their own content and display a sample of these contents on the outside of their cell membranes, using “Antigen” protein. Antigens cover the surface of cells displaying samples of proteins. As a key part of the human immune system, T-cells are designed to identify those displayed antigens by using proteins on their surface called as T-cell receptors(TCR). The T-cell receptors cover surface of the T-cells and bind to antigens. when a T-cell receptor fits with a viral antigen, it indicates the presence of a pathogen inside the cell.

Human blood has about one million T-cells in every milliliter (ml) [25]. Each one displays only one kind of T-cell receptor with the purpose that the variation in the repertoire will allow the identification of many kinds of diverse pathogens. When a given T-cell receptor does fit well with an antigen displayed protein, it signals the adaptive immune system to mount an immune response. In this case, three crucial steps occur: first, while still bound to the antigen displayed protein, the T-cell goes through a clonal expansion, making millions of copies of itself so that more T-cells can lock on to other infected cells with the same antigens display. Second, A subset of these T-cells changes phenotype and become killer T-cells. Those killer T-cells locate the virus-infected cells and puncture them, effectively killing them. Finally, a certain subset of the activated T-cells becomes memory T-cells, forming a standing army against future attacks from the same virus. This immunological memory is a critical component of vaccine development. [30] Figure 3 shows the steps a virus infects a host cell and how T-cells mount an immune system to kill a virus infected cell.

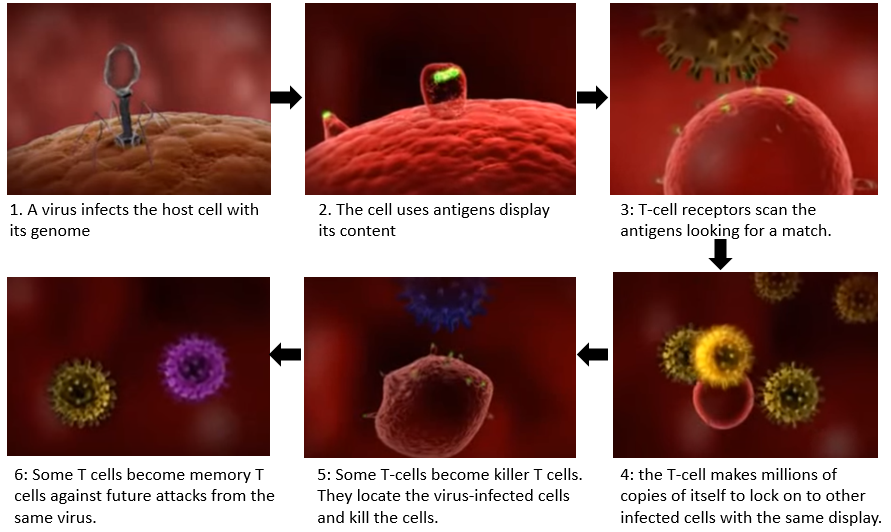


Figure 3: T-cells trigger an immune system to kill a cell that infected by a virus. [29]

Clearly, T-cells are critical in immune system for removing harmful cells. T-cell receptors bind to foreign proteins presented on harmful cells and lead to killing of the harmful cells. Unfortunately, T-cells often do not have a receptor that can bind to the proteins on cancer cells. This allows most cancers to escape killing by the T-cells and proliferate. In this case, T-cells needs antibodies’ help to identify uncovered antigens. Antibodies, which are produced by white blood cells called B cells, are large, Y-shaped proteins that recognizes antigens. Once a B cell’s antibodies have recognized an antigen, T-cells stimulate the B cell to produce large amounts of antibody specific to the pathogen being attacked so that the antibodies can mark the pathogen for T-cell recognition and destruction. Memory cells that are created by B cells, staying in the blood stream and produce mass amounts of antibodies for a specific antigen if the body is presented with it a second time. To cure cancer, specific antibody recognition of antigen is crucial in the body’s ability to combat a cancer cells. In other words, human body needs to use antibody mediated immune response to fight cancer.

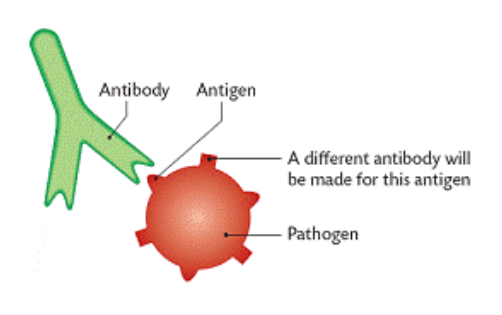


Figure 4: An antibody detects an antigen [35]

Immunotherapy strengthens a patient’s immune system to attack tumors by designing engineered T-cells in patients’ body to target and kill specific cancer cells. According to Leukemia & Lymphoma society, Immunotherapy is “A type of therapy that uses substances to stimulate or suppress the immune system to help the body fight cancer, infection, and other diseases. Some types of immunotherapy only target certain cells of the immune system. Others affect the immune system in a general way.” Two of the most promising new strategies in immunotherapy for fighting cancer, immune checkpoint inhibition [14] and chimeric antigen receptor T-cells [15], rely on the power of T-cells, endogenous or engineered, to recognize and specifically destroy cancerous cells.

***Immune checkpoint inhibition*** This treatment uses “checkpoint inhibitors” to activate an individual’s immune system to start an immune response to cancer cells. Studies [16] have shown that Immune checkpoints, also called inhibitory signaling pathways, may prevent the mounting of an immune response. Immune checkpoint inhibitors are drugs that prevent cancer cells from turning off T-cells. This allows T-cells to infiltrate a tumor and stop its growth. Several checkpoint inhibitors are currently FDA approved for the treatment of melanoma and lung cancer.

***Chimeric antigen receptor T-cells (CAR T-cell)*** This treatment collects patient’s own immune cells to treat their cancer.  The therapy draws patient’s blood to extract T-cells. Then, using a disarmed virus, T-cells are genetically engineered to produce receptors on their surface, called chimeric antigen receptors (CARs). These synthetic receptors allow the CAR T-cells to recognize and attach to an antigen, on tumor cells. Finally, the CAR T-cells will be infused into the patient’s body. Ideally, the engineered T-cells multiply in the patient's’ body and, with guidance from their engineered receptor, recognize and kill the cancer cells. Research [17] [18] has shown that CAR T-cell therapy works effectively in strengthening the patient’s immune system to attack tumors, especially for blood cancers. In 2017, two CAR T-cell therapies were approved by the Food and Drug Administration (FDA), one for the treatment of children with [acute lymphoblastic leukemia](https://www.cancer.gov/types/leukemia/patient/child-all-treatment-pdq) and the other for adults with advanced lymphomas. [19]

## Quantitative Multiplex Co-Immunoprecipitation Platform facilitates T-cells research

To facilitate immunotherapy research and gain a deep understanding of the dynamic molecular processes involved in T-cell receptor activation, many efforts have been invested in the last two decades, from small-scale experiments that measure interactions between a few proteins, to large-scale screens using high-throughput techniques such as yeast two-hybrid [9] and co-immunoprecipitation [13]. However, none of them models the molecular circuitry responsible for T-cell receptor(TCR) activation or suppression. Quantitative Multiplex Co-Immunoprecipitation (QMI) [1], which models molecular circuitry responsible for T-Cell receptor (TCR) activation or suppression of the TCR signalosome, is a novel network approach proposed by Dr. Stephen E.P. Smith, an assistant professor of pediatrics at the University of Washington, in 2015.

QMI is a software-based platform performing hundreds of immunoprecipitations from a small set of biomaterials and generating “network biosignatures” which reflect the T-cells activation state. The purpose of QMI is to analyze abnormal cells with protein-protein interactions (PPIs) and, ultimately, provide doctors with the data necessary for engineering T-Cells that can help patients’ immune systems fight abnormal cells.

QMI is a comprehensive analysis platform that generates PiSCES (proteins in shared complexes detected by exposed surface epitopes) biosignatures and can be used for discovering disease pathogenesis. With a 4-mm skin punch biopsy from control patients, QMI can generate PiSCES biosignatures by applying unsupervised hierarchical clustering, principal component analysis, an adaptive nonparametric with empirical cutoff analysis, and weighted correlation network analysis. QMI analysis can (1) distinguish patients from control groups, (2) detect autoreactive T-cell signaling in the autoimmune patients, and (3) generate a hypothesis regarding a disease-associated network signature [1].

Because QMI is a scalable multiplex approach that is of interest in disease pathogenesis, SEPS lab proposes to integrate it into two clinical cancer research groups (end-users) at two sites, the Mayo Clinic in Minnesota and Seattle Children’s Hospital in Washington State. SEPS lab plans to extensively validate the precision, accuracy, consistency, and stability of QMI data across user groups and work sites to demonstrate the robustness of QMI methodology. Specific goals include:

First, they intent to **validate and integrate the QMI platform into a CAR T-cell production pipeline**. Chimeric antigen receptor(CAR) T-cells are an engineered cell that show great promise in treating cancers. QMI platform has potential applications in CAR T-cell process, such as selecting the most promising chimeric antigen receptors for preclinical development based on the results of the QMI analysis, and analyzing CAR T-cells that have been recovered from blood after infusion into patients. Since Seattle Children’s research institute is a leader in CART cell therapy, SPES lab wants to validate the QMI method at the lab, to check consistent TCR signalosome signatures from CART cells, and to integrate the QMI system into the end users’ CART workflow.

Second, the lab aims to **validate and integrate the QMI platform into Immune checkpoint inhibition therapy**. Immune checkpoint inhibition therapy activates immune system to start an immune response toward cancer cells. Melanoma produces global immune suppression in late-stage patients, and responds to immune checkpoint therapy. Mayo clinic is a leader Melanoma treatment, and the Marchovic lab manages a biobank containing over 10 years of clinical materials, including viability freeze-downs of T-cells from patients treated with checkpoint inhibitor therapy. SPES lab proposes to validate the QMI method produces consistent TCR signalosome signatures from in primary melanoma patients’ cells, and integrate QMI platform into the end users’ workflow.

## A data analysis and virtualization tool for QMI platform

Successful completion of integration into end user labs will produce a validated standard operating procedure for QMI in cancer research, and inform the basic biochemical mechanisms of cancer pathogenesis and treatment. However, current QMI platform lacks usability and visualization that are essential for the validation and integration in the end-user labs. First, the QMI platform requires expert level of technical understanding to run three different programs: R, MATLAB, and high-level analysis in Excel. Second, current QMI platform lacks usability. It runs on command-line interfaces to execute different programs. Without detail information of how to implement QMI analysis, such as experimental procedures and controls, the complexity of the approach has prevented widespread adoption. A united interface and an automated process should be created to facilitate the integration into end-user labs.

Third, the QMI analysis lacks visualization. Since QMI only presents the final results of complex statistical workflows in a random order, users find it difficult to understand and analyze the results. Hence, users are unwilling to adopt the QMI technology without assurance that the technique will produce understandable, consistent, meaningful data. A visualized presentation for every step of experiments settings and analyses can leverage to produce consistent, meaningful data for users.

Given that QMI is a clinical process and requires customized setting for individual patients, it is important to automate the process and increase the usability and visualization. Therefore, we propose to develop a data analysis and vitalization program for QMI analysis. The benefits follow:

* It will produce a stream-lined bioinformatics workflow, integrating the MATLAB, R and Excel programs into a single easy-to-use program.
* It will provide a united user-friendly graphical user interface that makes operations simple and results understandable for end users.

Comparing to the existing QMI platform, our program will enable users to input biomaterial data and run various QMI analyses in a short time. It will also provide meaningful visualization to represent the QMI results associated with different types of T-cell activations.

# METHODOLOGY/SYSTEM DESIGN

In this section, we detail design aspects of the data analysis and virtualization program. We first describe the existing QMI architecture, following the overall architecture of the proposed system. We illustrate detailed implementation and designs choices.

## Requirements

Based on our communication with our client, Prof. Smith, who expects our project to provide usability and visualization to facilitate successful integration of the QMI platform. The user requirements are as follows:

1. This product should enable users to perform QMI analyses without a mediator.
2. This product should be easy to use by lab technicians and should be designed in a way that user errors are minimized.
3. This product shall provide meaningful visualization to help users understand QMI analysis and results.
4. Lab technicians shall be able to use all functions of this product after an initial training. After the training, the average number of errors made by experienced users shall not exceed two per day.
5. This product shall respond to users’ queries with fast responses. For example, when the users perform any analysis, the result must be presented within one minute.
6. This product shall be intuitive and self-explanatory.

## System Architecture

The existing QMI platform runs on three different programs: R, MATLAB, and manual Excel operations. Performing an QMI analysis requires the user to use command-line interfaces and manually input and manipulate data. Hence, an expert must work as a mediator to perform QMI analyses. The required back-and-forth communication for each process means that a typical analysis takes at least 30 minutes.

In order to improve the efficiency of the QMI analysis, we aim to achieve the following sub goals in our system: 1) A streamlined analysis process. Users with little or no knowledge of programming or statistics can perform the QMI analysis without a mediator. 2) Quick response. Users can perform one QMI analysis within 10 minutes. 3) Visual presentation of data and results. Users should easily configure setting parameters for the QMI analysis. We shall present analysis results logically with smooth transitions so that users can understand the results. Therefore, we customize programs that are used in the existing QMI platform and provide an intuitive graphical user interface(GUI) for simple operations and visualization. The steps with MATLAB programs will be replaced by a Java program due to MATLAB license cost and low maintainability. Figure 5 shows the components of the existing QMI platform and our program, and how users can perform a QMI analysis.

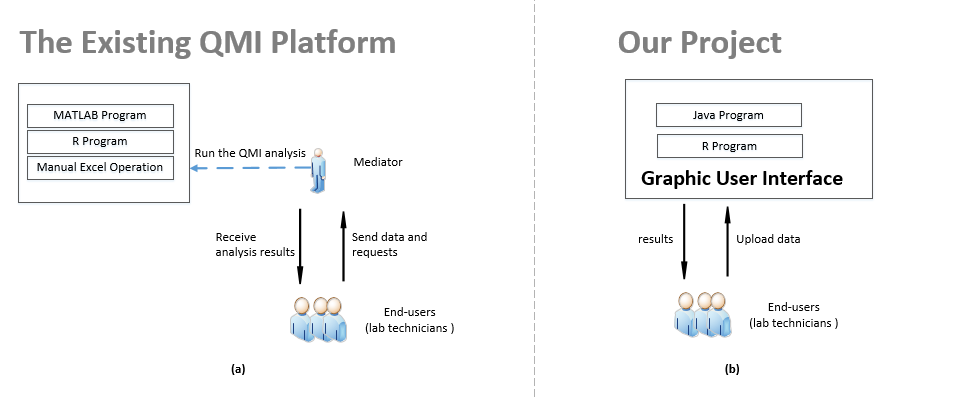


Figure 5: (a) shows the components of the existing QMI platform. Users need a mediator to perform an QMI analysis. (b) shows the components of our project. Three programs have been reduced to two programs, and users can upload and obtain results by themselves via a graphical user interface.

Figure 6 shows a data-flow diagram of the system. The system requires XML files that are generated from a Bioplex 200 instrument [1] as inputs. The machine performs an initial analysis on T-cells which are obtained and generated from control patients’ biopsy samples. T-cells are placed on an 8 by 12 plate, with 96 wells. To screen specific antibodies’ ability to capture and detect proteins from post-nuclear cell lysates, the lab technician adds antibodies to half of the T-cell preparations and then activate signaling. Data for antibodies reactions and T-cells activation are exported in XML format, and this XML files are inputs for the QMI analysis [35]. As a first step, the system parses XML files to acquire raw data for further processing. Then the system visualizes pairwise proteins’ raw data for each well in bead plates. Subsequently, the default analysis setting, which can be easily edited, is shown on the GUI based on number of the inputs. Finally, users perform a QMI analysis, with the results presented in the GUI and also saved into a data model, in case users need to check previous steps.

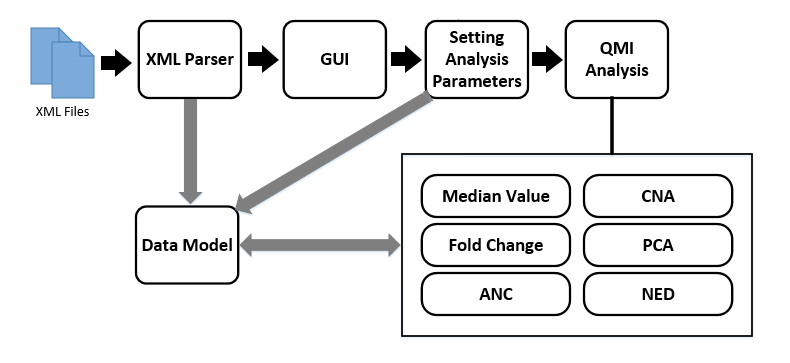


Figure 6: overall system data flow diagram

## Design Patterns

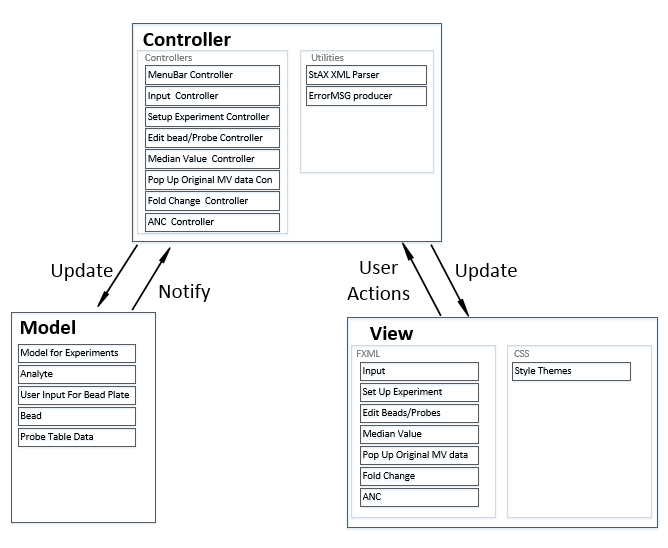
The program architecture applies well-established industry design practices. We use a Model-View-Controller framework to keep the layers of business logics and presentation separate. The Model is a black box representing the application and its data, and will typically be an n-tiered application. The “View” displays the data from the application and passes commands and data back to update the application. The “Controller” mediates between the “Model” and “View”, presenting data to the “View” and commands to the “Model” [30]. Since MVC supports rapid and parallel development and because modifications on one component does not affect others, MVC alleviates the complexity of large applications and makes developing and maintenance easy for developers. Figure 7 displays MVC implementation of the program. We also apply singleton, factory and builder design patterns in our program. 

Figure 7: Model-Controller-View framework

## Technologies and Tools

Below are the technologies and tools that are used in our program.

**JavaFx**

JavaFx is a software platform for creating and delivering desktop applications, as well as rich Internet applications (RIAs) that can run across a wide variety of devices. We chose JavaFX over other desktop user interface technologies such Java Swing because It is efficient to write a modular, clean, and maintainable code. First, it is fast and intuitive to build from scratch. It supports markup in “FXML” and Cascading Style Sheets (CSS) with a quick-start tool “Scene Builder”. FXML is an XML-based language which JavaFX uses to create the layout of screens. CSS is a stylesheet language that describes how elements should be rendered on screen. Scene Builder is an application which allows developers to drag and drop JavaFX UI components and mark interactions and CSS details. “FXML”, “CSS” and “Scene Builder” work together to makes implementation efficient. Second, JavaFX has better support for MVC Pattern than Java Swing. Using JavaFX developers can cleanly separate our work as views (FXML, CSS), models (Java, domain objects) and controllers (Java). On the contrary, its various components lack consistency makes MVC support in Java Swing is unsatisfying. Last, as stated by Oracle, JavaFX is the next step in their Java-based rich client strategy. Thus, we believe JavaFX will have better support and more features in the future.

**StAX XML parser**

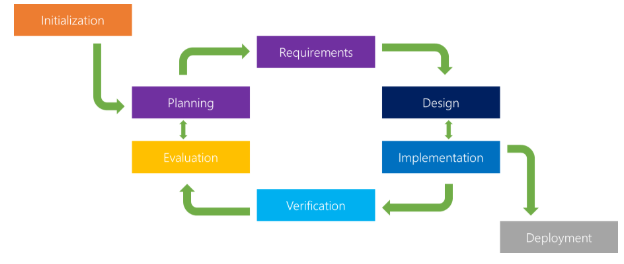
We chose the StAX XML parser to read inputs, because it is fast and memory efficient. XML (extensible Markup Language) is designed to store and transport data. Most current XML APIs fall into one of two broad classes: event-based APIs such as StAX or tree-based APIs such as DOM and JDOM. Although the tree-based APIs are user-friendly in implementation, they are less efficient in terms of memory usage because they read all the content in a file. StAX, as an advanced event-based parser with a streaming feature, uses much less memory space than a tree API because a client only gets XML data when it explicitly asks for it. In fact, StAX processes the document in small pieces, and frees program to wait for the entire document to be read. Additionally, it can read multiple documents simultaneously in one single thread. [32]

**R Language**

We retain existing R program for easy maintenance and flexible update. R is a highly-extensible program language providing a wide variety of statistical and graphical libraries. It is also a free software under the terms of the Free Software Foundation’s GNU General Public License. It compiles and runs on UNIX/Linux, Windows and MacOS [3]. Current QMI analysis run three complex statistical analyses with the R program, all of which are frequently updated. Therefore, we chose to maintain the R program, but call the R program from our unified system directly, so that the R program can be hidden but executable with our interface.

## Software Development Lifecycle Process

Our project utilizes the Iterative software development lifecycle process model. One challenge of our project is the fact that clients would not know the complete requirements of the final product from the beginning. To address this, our system should consider iteration of planning, requirement gathering and analysis, designing, building, and testing. Iterative model finds defects in an early stage and probes complete requirements from clients. In iterative model, we first present a high-level design for the project. Later, we built a skeleton version and evolved the product. Since we built and improved the product step by step, we can track the defects in the early stages and avoid the downward flow of the defects. Moreover, using the iterative model, we can obtain reliable and detailed feedbacks from clients. When presenting design and draft versions of our program to clients for their feedback, we are effectively asking them to give detailed requirements on how the product should work [33].

Figure 8: Iterative software development lifecycle process model [33]

## Analyses

The ultimate goal of the system is to process six analyses: “Median Value”, “Fold change”, “Adaptive nonparametric analysis(ANC)”, “Weighted correlation network analysis(CAN)”, “PCA”, and “Node-Edge Día(NED)”. So far, we have completed the first 3 analyses that were implemented with MATLAB. The remaining analyses will be implemented after discussing details with clients. Therefore, we describe the first three analyses below.

**Median Value**

Median value provides statistical evidence to validate the raw data in XML by calculating analytes’ median value of probes for each sample. For each well from a bead plate, data were first processed to eliminate doublets on the basis of the doublet discriminator intensity (>5000 and <25,000). Then we identify specific bead classes within the bead region used and pairs individual bead fluorescence measurements with their corresponding bead regions. If less than 50% of data are contaminated, the median will not give an arbitrarily large or small result. A median value exceeding 100 is judged as unusual [34].

**Fold Change**

Fold change is used to compare the expression of genes between two sets of arrays, e.g. case and control sets. Since fold change data exceeding a threshold indicates weakly detected or weakly changed PiSCES, it will be filtered in later analysis steps.

**Adaptive nonparametric analysis(ANC)**

ANC is customized for the empirical technical error in an experimental set and is used to define statistical significance, which means high-confidence in detecting autoreactive T-cell signaling. A two-sided Kolmogorov–Smirnov test (KS Test) with type 1 error of 0.05 is implemented to calculate P values and test hypotheses. If a P value for duplicate wells was less than 0.05, then those data will be excluded from further analysis because of presumed manual error.

## Measures

The following metrics will be used to evaluate the overall quality and usability of the GUI project:

* **Correctness**

This is correctness result of QMI analysis, meaning the accuracy of results returned by our project should be identical to the results provided by the current QMI program.

* **Effectiveness**

We use completion rate and task time to measure effectiveness. Completion rate = (number of tasks completed successfully /total number of tasks undertaken) \* 100%. Task time records how long it takes a user to complete a task in seconds and or minutes. Task time = End Time – Start Time. We target for a completion rate at 100%. For task time, we expect relatively simple analyses, with no more than 4 inputs, can be processed within 1 minute, and complicated ones should be finished within 5 minutes.

* **User Satisfaction**

Since a good user experience is essential for users to adopt the QMI platform in their workflow, we should have end-users to perform task tests. For this testing, we will ask participants to answer a few questions about their impression of the overall ease of use. However, since our project is incomplete at this point, we will postpone the survey until a future phase.

# RESULTS

In this section, we first present the results of comparing our program with the existing program. Then we evaluate our program using metrics discussed in Methods chapter.

## Running an QMI analysis

**Dataset**

We used 8 XML data files for input data as shown in figure 9. These files are generated by a Bioplex 200 instrument and provided by our client.

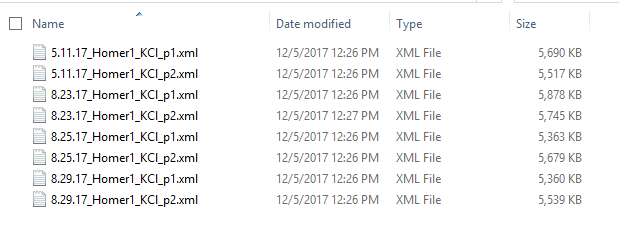
We used 8 XML data files as input

Figure 9: Input data to run a QMI analysis.

**Steps**

As shown in figure 10, there are 4 steps to run a QMI analysis: 1) Parameter setup with given inputs, 2) Median Value analysis, 3) Fold change analysis, and 4) ANC analysis. The remaining analysis steps, “Weighted correlation network analysis(CAN)”, “PCA”, and “Node-Edge Día(NED)”, will be implemented in the future work. Our program provides a flexibility, in that it allows the user to change input setting at any time during any process whenever necessary. Also, a user can switch between each analysis to view analysis results, i.e. personalize a order by rearranging the default order.

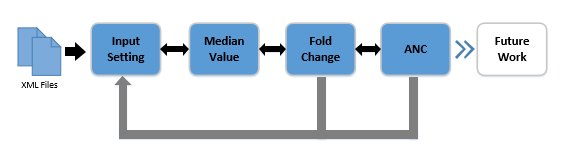


Figure 10: Steps to run QMI analysis using our project

Figure 11 shows the graphical user interface for the QMI analysis with a panel view for each process. In the “Input” panel, a user inputs source data and sets up parameters by clicking a “upload” button on Figure 11 (a) and selecting files. Parameters are automatically generated initially. Input parameters include number of experiments, xml files for each experiment, analytes, number of samples, number of replicas, number of probes, and probe names. These default settings can be changed in seconds. Moreover, our program provides visualization for bead plates setting. Through this visualization, in which plates represent wells and colors indicating probes, users can easily validate the experimental setup for the QMI analysis. Then, a user can move on to Median value figure 11 (b), fold change figure 11 (c) and ANC analysis figure 11 (d) analyses by changing the corresponding panel. Then, the user selects a sample or a sample combination to view results. The results are color-coded to highlight outstanding data.

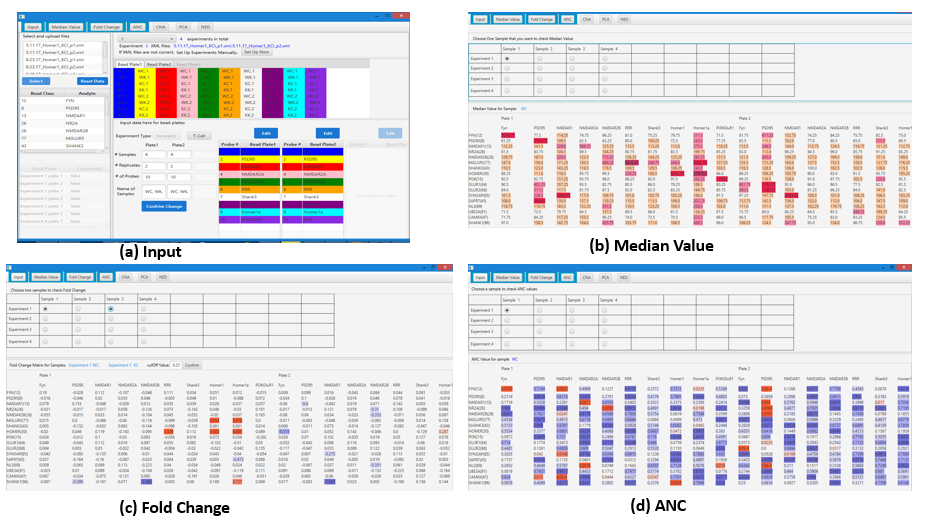


Figure 11: running QMI analysis using our project

In contrary to our project, the current QMI platform lacks flexibility and the capability of user interaction. For example, it requires users to change parameters directly in the MATLAB code to setup, and the process cannot be modified once the analysis starts. Additionally, it fails to organize the results for easy visualization, since all results are provided as a series of figures without organized layouts.

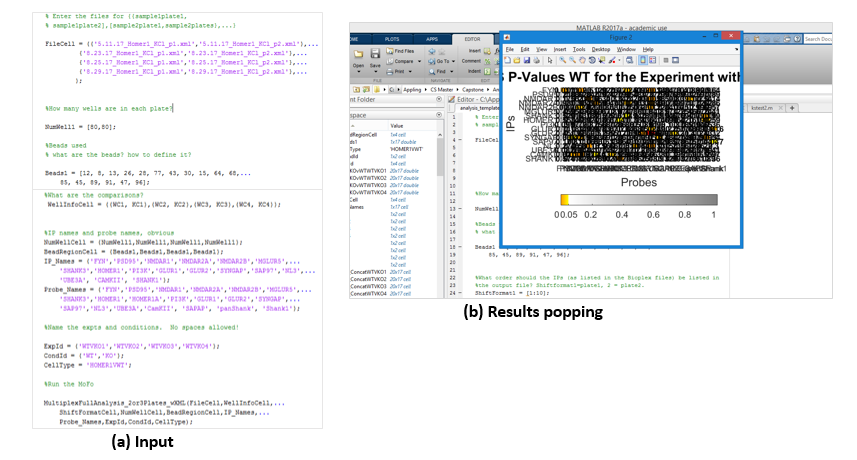


Figure 12: running QMI analysis using the existing QMI platform

## Evaluation

Here we present the evaluation results using metrics discussed in the Methods chapter.

**Correctness**. We conducted 4 set QMI analyses to check whether results are identical to those obtained by using the existing QMI platform. The results are 100% correct.

|  |  |  |
| --- | --- | --- |
| QMI analysis | Setting parameters | Correct? |
| 1 | 1 experiment with 2 XML files. | Y |
| 2 | 2 experiments with 4 XML files. | Y |
| 3 | 3 experiments with 6 XML files. | Y |
| 4 | 4 experiments with 8 XML files. | Y |

Figure 10: 4 QMI analyses setting to evaluate correctness

**Effectiveness.** We conducted 4 set QMI analyses with unusual setting to check whether our project able to produce results. The test analyses all passed. The completion rate was 100%. The task times, ranging from less than 1 minute to less than 3 minutes, satisfy users requirements. On the contrary, when used the existing QMI Platform, only 50% test cases passed and each task takes at least 4 minutes. Our program greatly improved the effectiveness of QMI Platform for users.

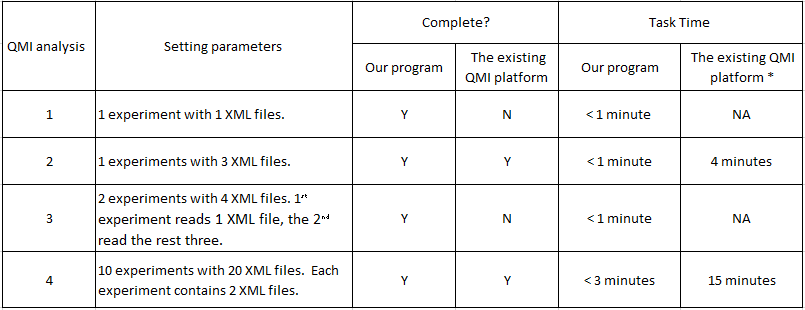


Figure 11: 4 QMI analyses setting to evaluate effectiveness.

\* Task time for the existing QMI platform was based on operations by an expert.

# CONCLUSION & FUTURE WORKS

Over a period of 10 months, to facilitate the integration of QMI platform into end user labs, we have implemented a data analysis and virtualization system to replace the existing QMI platform using Java and JavaFX. We have built a graphical user interface (GUI) based QMI analysis software program, which provides a streamlined workflow, and an intuitive and interactive interface by integrating three separate analytic programs into a single program. We implemented a Model-View-Control framework, which takes into account compatibility with existing platform, code implementation, testing, and communication with clients. We conducted user-centered design for our product to provide an intuitive, efficient and user-friendly experience for end users. We adopted the iterative development model to avoid the downward flow of the defects and to effectively probe requirements from clients. In the end, our program was evaluated using correct rate, completion rate, and task time. The results of these metrics indicate that our project scores high in correctness and effectiveness.

The existing QMI platform lacks the usability and visualization that are required by end user labs, problem which our project addressed. Our program enables users to input biomaterial data and run various QMI analyses efficient and fast. It also provided meaningful visualization to represent the QMI results associated with different types of T-cell activations. We believe our project will build users’ confidence in adopting QMI analysis and will have a significant impact on immunotherapy research by efficiently providing meaningful data about T-Cell receptor (TCR) activation and suppression.

In the future, we should continue on implementing “CAN”, “PAC” and “NED” analyses. After that, to assure a user-friendly experience, we need to conduct usability evaluation to gather comments and suggestions from users for improvements. Last, we shall install our project into end user labs, prepare technical documentation for future maintenance employees, and design and arrange trainings for end users.

**Acknowledgements**

I would like to thank professor Wooyoung Kim for encouraging me and guiding me through the project. I would like to thank Dr. Stephen E.P. Smith for providing requirements and explaining knowledge of QMI Platform necessary for our project. I would like to thank professor Yang Peng, professor Geetha Thamilarasu and professor Nancy Kool for valuable suggestions and proofreading on the project paper and the project. Moreover, I would like to thank my family and friends for all of their supports and encouragements throughout my master’s program. I would not have accomplished this project without the guidance and encouragement from any of these people.

**References**

1. Smith, S. E., Neier, S. C., Reed, B. K., Davis, T. R., Sinnwell, J. P., Eckel-Passow, J. E.& Neuhauser, C. (2016). Multiplex matrix network analysis of protein complexes in the human TCR signalosome. Sci. Signal., 9(439), rs7-rs7.
2. [1] Smith, S. E., Neier, S. C., Reed, B. K., Davis, T. R., Sinnwell, J. P., Eckel-Passow, J. E. & Neuhauser, C. (2016). Supplementary Materials for Multiplex matrix network analysis of protein complexes in the human TCR signalosome. Sci. Signal., 9(439), rs7-rs7.
3. The R Project for Statistical Computing. https://www.r-project.org/about.html
4. Jones, S., & Thornton, J. M. (1995). Protein-protein interactions: a review of protein dimer structures. Progress in biophysics and molecular biology, 63(1), 31-65.
5. Douglass, A. D., & Vale, R. D. (2005). Single-molecule microscopy reveals plasma membrane microdomains created by protein-protein networks that exclude or trap signaling molecules in T-cells. Cell, 121(6), 937-950.
6. Wiederschain, G. Y. (2006). Protein-protein interactions. A molecular cloning manual. Biochemistry (Moscow), 71(6), 697-697.
7. Khatib-Shahidi, S., Andersson, M., Herman, J. L., Gillespie, T. A., & Caprioli, R. M. (2006). Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry. Analytical chemistry, 78(18), 6448-6456.
8. Pemble, S., Schroeder, K. R., Spencer, S. R., Meyer, D. J., Hallier, E., Bolt, H. M., & Taylor, J. B. (1994). Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. Biochemical Journal, 300(Pt 1), 271.
9. Sato, T., Hanada, M., Bodrug, S., Irie, S., Iwama, N., Boise, L. H., & Wang, H. G. (1994). Interactions among members of the Bcl-2 protein family analyzed with a yeast two-hybrid system. Proceedings of the National Academy of Sciences, 91(20), 9238-9242.
10. Sekar, R. B., & Periasamy, A. (2003). Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. The Journal of cell biology, 160(5), 629-633.
11. The SPES Lab. http://faculty.washington.edu/seps/SEPS\_Lab/The\_SEPS\_Lab.html
12. Mahe, E., Pugh, T., & Kamel-Reid, S. (2017). T-cell clonality assessment: past, present and future. Journal of clinical pathology, jclinpath-2017.
13. Kuroda, K., Kato, M., Mima, J., & Ueda, M. (2006). Systems for the detection and analysis of protein–protein interactions. Applied microbiology and biotechnology, 71(2), 127-136.
14. Topalian, S. L., Drake, C. G., & Pardoll, D. M. (2015). Immune checkpoint blockade: a common denominator approach to cancer therapy. Cancer cell, 27(4), 450-461.
15. Maude, S. L., Frey, N., Shaw, P. A., Aplenc, R., Barrett, D. M., Bunin, N. J., & Mahnke, Y. D. (2014). Chimeric antigen receptor T-cells for sustained remissions in leukemia. New England Journal of Medicine, 371(16), 1507-1517.
16. Dine, J., Gordon, R., Shames, Y., Kasler, M. K., & Barton-Burke, M. (2017). Immune checkpoint inhibitors: an innovation in immunotherapy for the treatment and management of patients with cancer. Asia-Pacific journal of oncology nursing, 4(2), 127.
17. Maude, S. L., Teachey, D. T., Porter, D. L., & Grupp, S. A. (2015). CD19-targeted chimeric antigen receptor T-cell therapy for acute lymphoblastic leukemia. Blood, 125(26), 4017-4023.
18. Lee, D. W., Kochenderfer, J. N., Stetler-Stevenson, M., Cui, Y. K., Delbrook, C., Feldman, S. A., & Steinberg, S. M. (2015). T-cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial. The Lancet, 385(9967), 517-528.
19. Porter, D. L., Hwang, W. T., Frey, N. V., Lacey, S. F., Shaw, P. A., Loren, A. W., & Ambrose, D. (2015). Chimeric antigen receptor T-cells persist and induce sustained remissions in relapsed refractory chronic lymphocytic leukemia. Science translational medicine, 7(303), 303ra139-303ra139.
20. Yeger-Lotem, E., & Sharan, R. (2015). Human protein interaction networks across tissues and diseases. Frontiers in genetics, 6, 257.
21. Rosenberg, S. A., Packard, B. S., Aebersold, P. M., Solomon, D., Topalian, S. L., Toy, S. T., & Simpson, C. (1988). Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. New England Journal of Medicine, 319(25), 1676-1680.
22. Rosenberg, S. A., Aebersold, P., Cornetta, K., Kasid, A., Morgan, R. A., Moen, R., & Merino, M. J. (1990). Gene transfer into humans—immunotherapy of patients with advanced melanoma, using tumor-infiltrating lymphocytes modified by retroviral gene transduction. New England Journal of Medicine, 323(9), 570-578.
23. Topalian, S. L., Solomon, D., Avis, F. P., Chang, A. E., Freerksen, D. L., Linehan, W. M., & Simon, P. (1988). Immunotherapy of patients with advanced cancer using tumor-infiltrating lymphocytes and recombinant interleukin-2: a pilot study. Journal of Clinical Oncology, 6(5), 839-853.
24. Lymphatic system and immunity. https://www2.estrellamountain.edu/faculty/farabee/biobk/BioBookIMMUN.html
25. Sherwood, L. (2011). Fundamentals of human physiology. Cengage Learning
26. Griffiths, A. J. (2005). An introduction to genetic analysis. Macmillan.
27. Sudhakar, A. (2009). History of cancer, ancient and modern treatment methods. Journal of cancer science & therapy, 1(2), 1.
28. Janeway, C. A. (2001). How the immune system works to protect the host from infection: a personal view. Proceedings of the National Academy of Sciences, 98(13), 7461-7468.
29. How T-cells work. https://www.youtube.com/watch?v=kIxmiTuRydw
30. Krasner, G. E., & Pope, S. T. (1988). A description of the model-view-controller user interface paradigm in the smalltalk-80 system. Journal of object oriented programming, 1(3), 26-49.
31. McDonough, J. E. (2017). Singleton Design Pattern. In Object-Oriented Design with ABAP (pp. 137-145). Apress, Berkeley, CA.
32. Kataoka, K. (2006). U.S. Patent No. 7,013,425. Washington, DC: U.S. Patent and Trademark Office.
33. Iterative Model: What Is It and When Should You Use It. https://airbrake.io/blog/sdlc/iterative-model
34. Median. https://en.wikipedia.org/wiki/Median
35. The Human Defense System. http://leavingbio.net/human-defence-system/