Utah Population based genetic and clinical feature in Colorectal cancer

BY HYOJOON PARK & SEYOUN BYUN

BASIC INFORMATION

Link to the project: https://github.com/seyoun209/dataviscourse-pr-coloncancer
Hyojoon Park u1266489@utah.edu u1266489
Seyoun Byun u0693520@utah.edu u0693520

BACKGROUND AND MOTIVATION

Colorectal cancer is the third most common cancer diagnosed yearly, in both men and women, in the United States and the second leading cause of cancer-related deaths when men and women are combined. Colorectal cancer treatment is beneficial and reducing the number of incidences when removing the colon polyps. Also, early diagnosis and survival are better with detecting the polyps in the colon.

Interestingly, the asserted pedigree and twin studies indicate that 20-30% of colon cancer cases appear to arise in the inherited susceptibility. Also, 3-5% of colon cancer occurs in inherited syndrome. However, the risk of adenomatous polyps in men and women or different ages concerning the family history of colon cancer cases is not studied well.

Association studies reported genetic variants and exposure risk factors, including BMI, smoking, exercise, alcohol consumption, NASID, and hormone menopause.

This project will use the seven families with several patients (n=198)'s genetic information. We initiate genetic information to compared familial information. We will respectably sort out all exposure risk with ascending/descending method to understand which exposure risk is most relatively related to colon cancer. Lastly, we will try to understand the polyp size with the relationship with the family's inheritance. Therefore, in this visualization, we evaluated the polyp size with the location and the clinical factors in huge extended families with a strong family history representing the familial high-risk colorectal cancer classification.

PROJECT OBJECTIVE

The objective is to utilize interactive plots to help viewers explore relations between polyps sizes and other biological and family-related factors. The plots allow users to sort and filter data such that the relations can be viewed from different perspectives. This would allow users draw conclusions on how the various factors affect the incidence of colon cancer.

DATA

The Utah population database (UPDB) was used to identify the seven families. Colorectal cancer cases in the families were contacted by the Utah Cancer Registry through mail requesting them, or their next of kin, permission to be contacted by the study. In total, there are seven large kindreds with multiple colorectal cancer cases included in this study. The medical records were obtained on colorectal cancer cases. Published guidelines evaluated adenomatous and hamartomata’s polyposis syndromes.

Due to the IRB, the data is not appropriated to provided.

DATA PROCESSING

We obtain the raw data from UPDB. For effective display of information, we clean up the raw data such that data categories with many missing data are discarded from the plots. We narrow down the data exposure risks for display as the following:

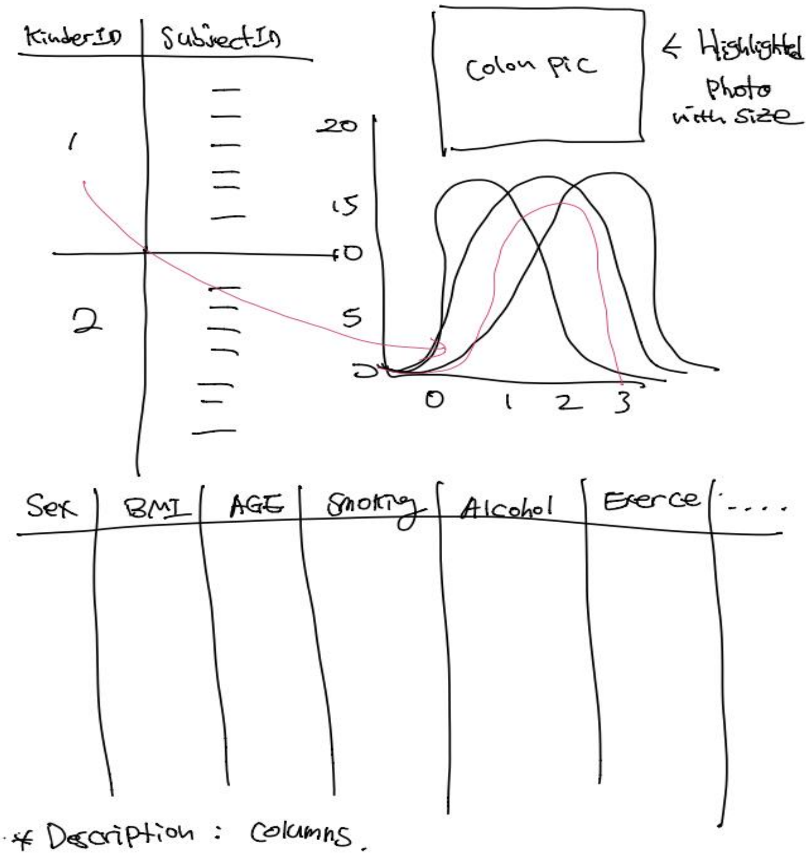
- Kinder ID, Subject ID, Sex, BMI, Age, Smoking, Alcohol, Exercise, NASID, HRT

In the second screen, we display the position and size of polyposis on the image of colon with the following information:

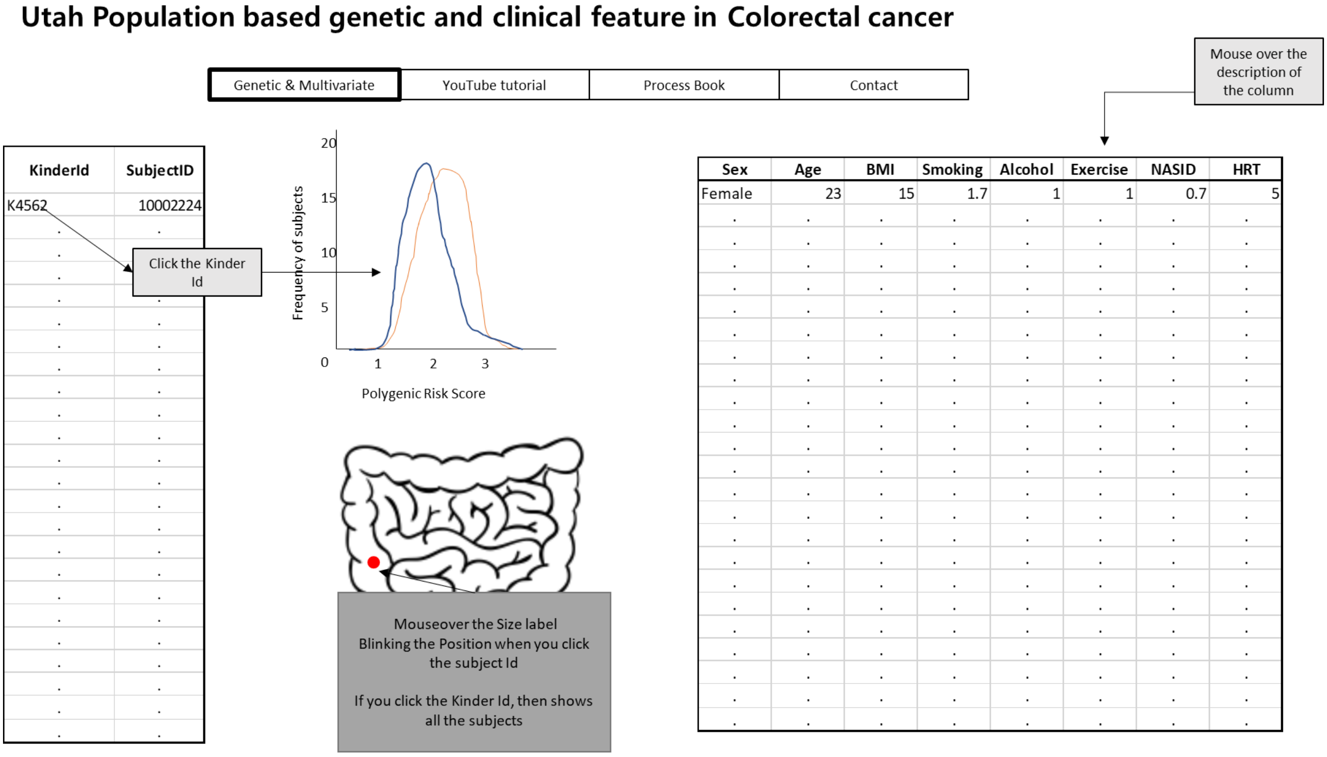
- Kinder ID, Subject ID, Site, Polytype, Size of Polyp.

VISUALIZATION DESIGN

Brainstorming



Initial Design 1 (Not using): We are not using this design because it might confuse the audience. Since our data is relatively complicated due to multiple variants, it does not focus on our motivation, which compares each kinder Ids. We have to use lots of hoovering tooltips that may cause the data even more complicated and not delivering well with our purpose. We are concerned that this is too much, just a data table instead of visualizing.



Utah Population based genetic and clinical feature in Colorectal cancer

Genetic & Multivariate | YouTube tutorial | Process Book | Contact

Mouse over the description of the column

KinderId	SubjectID
K4562	10002224
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.

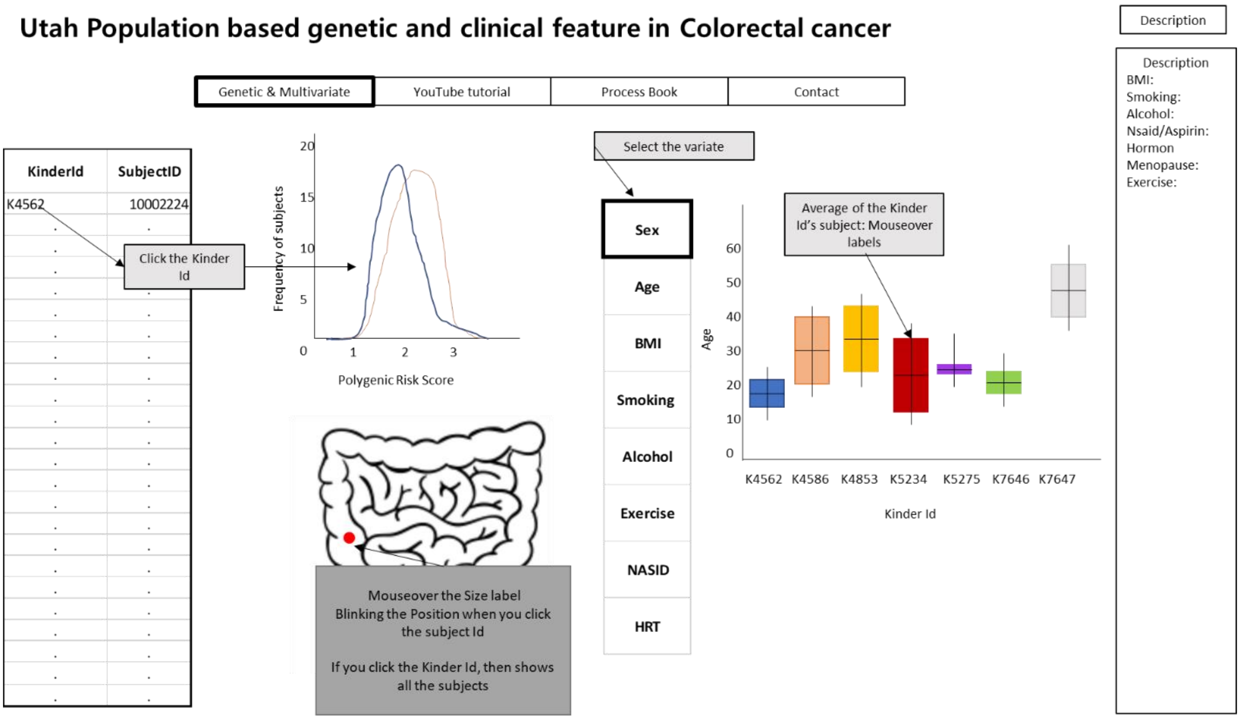
Click the Kinder Id

Mouseover the Size label
Blinking the Position when you click the subject Id

If you click the Kinder Id, then shows all the subjects

Sex	Age	BMI	Smoking	Alcohol	Exercise	NASID	HRT
Female	23	15	1.7	1	1	0.7	5
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

Initial Design 2 (Not Using): We are not using this design because all the data is compacted in one page. Every information will be average, including density plot and also all the multivariate features. In that case, we are not able to track down individual subject value. Nevertheless, we were trying to see the overall average of each family may be beneficial. We decided to split two slides to better visualize all the data by focusing on clinical and genetic information.



Utah Population based genetic and clinical feature in Colorectal cancer

Description

Genetic & Multivariate | YouTube tutorial | Process Book | Contact

KinderId	SubjectID
K4562	10002224
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.

Click the Kinder Id

Mouseover the Size label
Blinking the Position when you click the subject Id

If you click the Kinder Id, then shows all the subjects

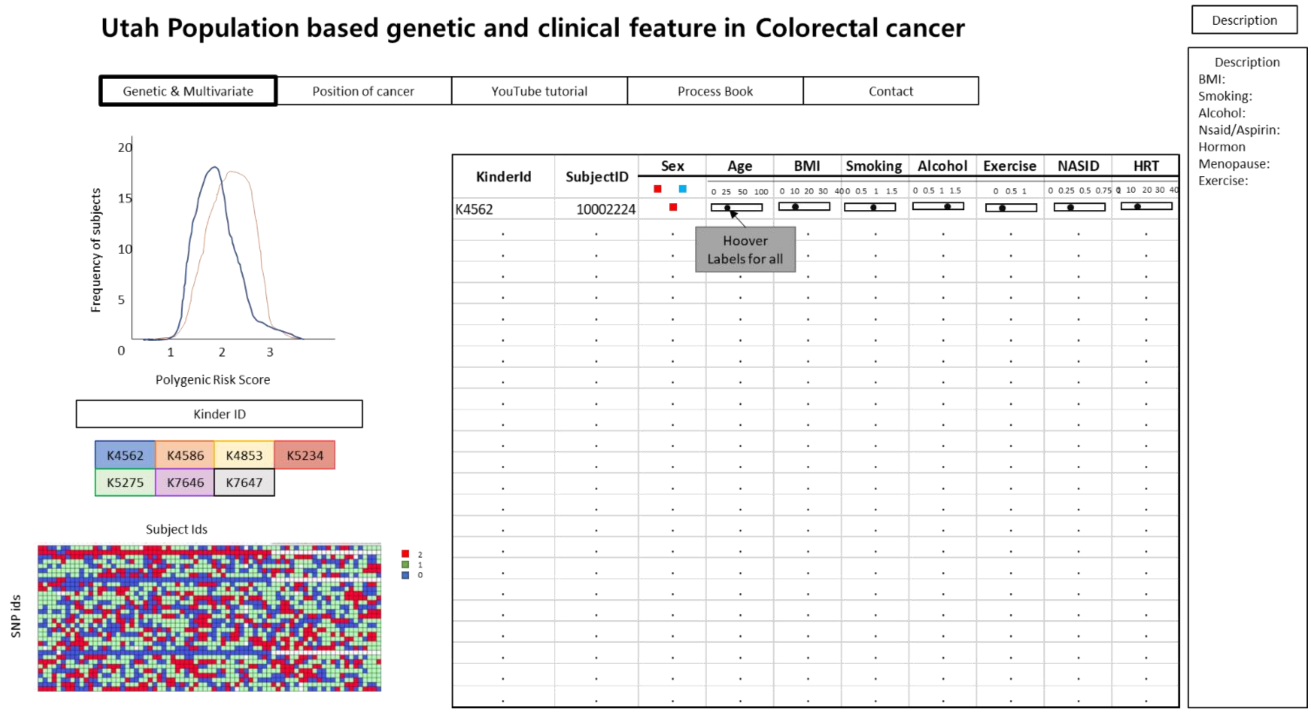
Select the variate

Average of the Kinder Id's subject: Mouseover labels

Description
BMI:
Smoking:
Alcohol:
Nsaid/Aspirin:
Hormon
Menopause:
Exercise:

Initial Design 3 (Using): We choose this design for our visualization. Since we have multiple clinical information, we decided to have two different slides to visualize our data.

Screen 1: We decided to combine the kinder-ID and subject-ID in one table with the clinical information, so the table is less complicated. We can also have a bigger density plot with the colored button of the kinder-ID to compare each family. Using the heatmap may show the genetic information better depending on the families. Finally, we added the bar with the circle. Thus, we know the numbers of each feature may be intuitive at first sight.



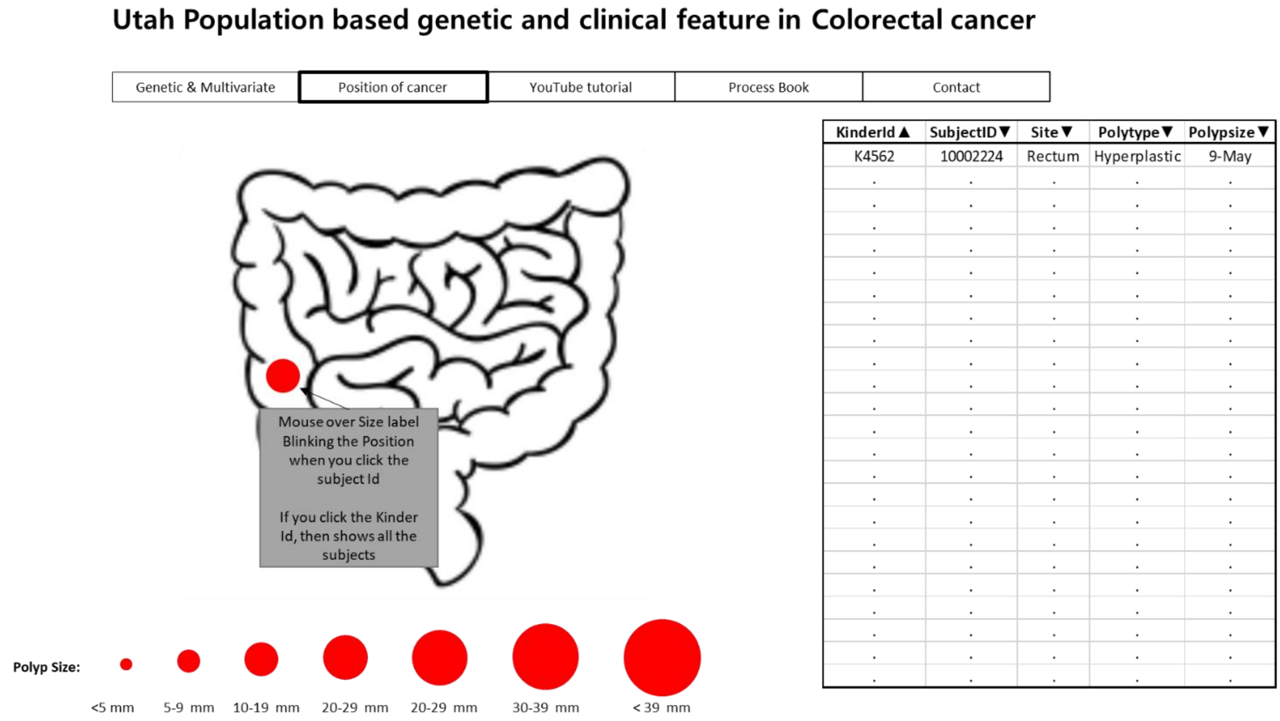
Utah Population based genetic and clinical feature in Colorectal cancer

KinderId	SubjectID	Sex	Age	BMI	Smoking	Alcohol	Exercise	NASID	HRT
		🟥 🟦	0 25 50 100	0 10 20 30 40	0 0.5 1 1.5	0 0.5 1 1.5	0 0.5 1	0 0.25 0.5 0.75 1	0 10 20 30 40
K4562	10002224	🟥	●	●	●	●	●	●	●
.	.	.		.	.	.	.	.	.
.	.	.		.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

Description
BMI:
Smoking:
Alcohol:
Nsaid/Aspirin:
Hormon
Menopause:
Exercise:

Screen 2: This is polyposis information of the colon cancer with each sample. Polyp information is another critical factor in our data, so; we decided to have a second slide to visualize better. We decided to have a table with sorting information for each feature. Using the circle size shows the extent of polyps as well as a location in the colon. We can interact with the kinder-ID and subject-ID using the klick method to better visualize the size and area.

We have all the value in the table so, less complication to imagine our data.



Mouse over Size label Blinking the Position when you click the subject Id

If you click the Kinder Id, then shows all the subjects

KinderId▲	SubjectID▼	Site▼	Polytype▼	Polypsize▼
K4562	10002224	Rectum	Hyperplastic	9-May
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Must-Have Features

- Table of the clinical records with multiple features including sex, age, BMI, smoking, alcohol, exercise, NASID, and HRT, will be provided with the sorting. All the value will be hoovered with the mouseover.
- The density plot with the seven kindred ID will be provided by clicking. Thus, two or multiple lines will be compared.
- The description will be on the side, and the Description clicking icon can remove it.
- The polyp information table, including kinder-ID, subject-ID, site, polyp type, and polyp size, will be provided with the sorting. All the value will be hoovered with the mouseover.

Optional Features

- Provided the heatmap with the genetic information for the details for the density plot.
- Provided the colon picture with the highlighted polysized and the location with interacting subject-ID and/or Kinder-ID.

Project Schedule

- 11-06-2020: complete setting up the layouts
- 11-13-2020: complete data parsing and setup for display
- 11-20-2020: complete basic displays of parsed data
- 11-27-2020: add interactive features
- 12-02-2020: final check and submit