현장연구 E 보고서

합성생물학 재현성에 관한 연구

김승화

21 June, 2021

Table of Contents

# 연구 개요

합성생물학 발전의 원동력이 되었던 iGEM 사례 분석을 통해 합성생물학이라는 개념을 정립하고 사용된 부품,회로들의 정량적 데이터 수집 및 재현성을 분석하고 Rmarkdown/Rstudio 활용능력을 학습해본다.

## 연구의 필요성 및 목표

합성생물학이란 생물학에 공학의 개념을 도입하여 복잡한 생명현상의 원리를 실제 만들어가며 이해하려는 시도이다. 공학에서의 대부분의 결과물이 그렇듯 합성생물학연구는 실제 문제에 대한 해결책을 제시하기 위한 목적을 가지고 있다. 이를 R mark down tool과 iGEM website를 통해 실습 및 연구를 진행하여본다.

# 연구 방법

## 연구 방법 1

**R tool 다루기**

1. Creating Git hub page

2-1. Create a new Github repository with pages 2-2. Create a new project in R studio 2-3. Connect local project to Github repository 2-4. Github page 생성 2-5. New page 생성 2-6. 로컬, 리모트 저장소에 commit 및 push 2-7. 웹 페이지 확인

3-1. page 완성 3-2. iGEM part Rmarkdown page 만들기 3-3. iGEM team table

no <- c(1,2,3,4,5)  
  
team\_name <- c("GO\_Paris-Saclay", "BOKU-Vienna", "BGU-Israel", "Hannover", "KAIT-Japan")  
project\_title <- c("HuGenesS", "Phangel-Taking Phage Therapy Ahead", "WIPEOUT-Wipes out wet wipes of the environment", "InToSens Development of an Inflammatory Toxin Sensor for detecting implant associated inflammations","E.coli that Create a Creative Environment")  
  
project\_year <- c(2020,2020,2020,2020,2020)  
  
wiki\_page <- c("https://2020.igem.org/Team:GO\_Paris-Saclay", "https://2020.igem.org/Team:BOKU-Vienna", "https://2020.igem.org/Team:BGU-Israelndo5a", "https://2020.igem.org/Team:Hannover", "https://2020.igem.org/Team:KAIT\_Japan")  
  
igem\_team <- data.frame(no,   
 team\_name,  
 project\_title,  
 project\_year,  
 wiki\_page)  
#knitr::kable(std, format = "markdown")  
   
knitr::kable(igem\_team)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| no | team\_name | project\_title | project\_year | wiki\_page |
| 1 | GO\_Paris-Saclay | HuGenesS | 2020 | <https://2020.igem.org/Team:GO_Paris-Saclay> |
| 2 | BOKU-Vienna | Phangel-Taking Phage Therapy Ahead | 2020 | <https://2020.igem.org/Team:BOKU-Vienna> |
| 3 | BGU-Israel | WIPEOUT-Wipes out wet wipes of the environment | 2020 | <https://2020.igem.org/Team:BGU-Israelndo5a> |
| 4 | Hannover | InToSens Development of an Inflammatory Toxin Sensor for detecting implant associated inflammations | 2020 | <https://2020.igem.org/Team:Hannover> |
| 5 | KAIT-Japan | E.coli that Create a Creative Environment | 2020 | <https://2020.igem.org/Team:KAIT_Japan> |

3-4. iGEM part table

no<- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30)  
  
part\_id <- c("BBa\_K2406020","BBa\_B1001"," BBa\_K3514001","BBa\_K3514002","BBa\_K3514003","BBa\_B0040","BBa\_K3514004","BBa\_K3338000","BBa\_K3338001","BBa\_K3338002","BBa\_K3338003","BBa\_K3338004","BBa\_K3338005","BBa\_K3338006","BBa\_K3338007","BBa\_K3338008","BBa\_K3338010","BBa\_K3338011","BBa\_K3338012","BBa\_K3338013","BBa\_K3338015","BBa\_K3338014","BBa\_K3338016","BBa\_K3338017","BBa\_K3338018","BBa\_K3338019","BBa\_K3338021","BBa\_K3338022","BBa\_K3338023","BBa\_K3338020")  
  
part\_name <- c("T7-LacO Promoter","Terminator(artifical, small, %T~=90)","Protein Exo of the Lambda Red System, degrades dsDNA into ssDNA, leaves 3' ends","Protein Beta of the Lambda Red System","Protein Gam of the Lambda Red System","Spacer.1 (generic)","Lambda Red System","Iron transporter MagA from Magnetospirillum magneticum","Human codon optimized Gaussia luciferase hGLuc","Synthetic promoter\_2 with NF-κB and AP1 binding sites","P2A self-cleaving peptide without GSG-linker","Internal ribosome entry site (IRES) for use in eukaryotic cells","Interleukin-6 Promoter mutagenized (IL-6 Pmut)","EGFP","Synthetic promoter\_1 with NF-κB and AP1 binding sites","Interleukin-6 Promoter (IL-6 P)","IL-6 P-MagA-P2A-hGLuc (Inflammatory Toxin Sensor)","CMV-EGFP-MagA-P2A-hGLuc","CMV-EGFP-MagA","CMV-EGFP-P2A-mCherry","CMV-EGFP-MagA-IRES-hGLuc","CMV-EGFP-IRES-mCherry","CMV-hGLuc-FLAG","CMV-hGLuc","CMV-MagA-FLAG","IL-6 Pmut-hGLuc","SynthP\_1-hGLuc","SynthP\_2-hGLuc","IL-6 Pmut-MagA-P2A-hGLuc","pEGFP-C2")  
  
part\_type <- c("promoter","terminator", "protein","protein","protein","Spacer","Lambda system","Reporter","Reporter","Regulatory","Coding","RBS","Regulatory","Reporter","Regulatory","Regulatory","Composite","Composite","Composite","Composite","Composite","Composite","Composite","Composite","Composite","Composite","Composite","Composite","Plasmid","Plasmid")  
  
team\_id <- c(2,2,2,2,2,2,2,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4)  
  
igem\_part <- data.frame(no,  
 part\_id,  
 part\_name,  
 part\_type,  
 team\_id)  
knitr::kable(igem\_part)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| no | part\_id | part\_name | part\_type | team\_id |
| 1 | BBa\_K2406020 | T7-LacO Promoter | promoter | 2 |
| 2 | BBa\_B1001 | Terminator(artifical, small, %T~=90) | terminator | 2 |
| 3 | BBa\_K3514001 | Protein Exo of the Lambda Red System, degrades dsDNA into ssDNA, leaves 3’ ends | protein | 2 |
| 4 | BBa\_K3514002 | Protein Beta of the Lambda Red System | protein | 2 |
| 5 | BBa\_K3514003 | Protein Gam of the Lambda Red System | protein | 2 |
| 6 | BBa\_B0040 | Spacer.1 (generic) | Spacer | 2 |
| 7 | BBa\_K3514004 | Lambda Red System | Lambda system | 2 |
| 8 | BBa\_K3338000 | Iron transporter MagA from Magnetospirillum magneticum | Reporter | 4 |
| 9 | BBa\_K3338001 | Human codon optimized Gaussia luciferase hGLuc | Reporter | 4 |
| 10 | BBa\_K3338002 | Synthetic promoter\_2 with NF-κB and AP1 binding sites | Regulatory | 4 |
| 11 | BBa\_K3338003 | P2A self-cleaving peptide without GSG-linker | Coding | 4 |
| 12 | BBa\_K3338004 | Internal ribosome entry site (IRES) for use in eukaryotic cells | RBS | 4 |
| 13 | BBa\_K3338005 | Interleukin-6 Promoter mutagenized (IL-6 Pmut) | Regulatory | 4 |
| 14 | BBa\_K3338006 | EGFP | Reporter | 4 |
| 15 | BBa\_K3338007 | Synthetic promoter\_1 with NF-κB and AP1 binding sites | Regulatory | 4 |
| 16 | BBa\_K3338008 | Interleukin-6 Promoter (IL-6 P) | Regulatory | 4 |
| 17 | BBa\_K3338010 | IL-6 P-MagA-P2A-hGLuc (Inflammatory Toxin Sensor) | Composite | 4 |
| 18 | BBa\_K3338011 | CMV-EGFP-MagA-P2A-hGLuc | Composite | 4 |
| 19 | BBa\_K3338012 | CMV-EGFP-MagA | Composite | 4 |
| 20 | BBa\_K3338013 | CMV-EGFP-P2A-mCherry | Composite | 4 |
| 21 | BBa\_K3338015 | CMV-EGFP-MagA-IRES-hGLuc | Composite | 4 |
| 22 | BBa\_K3338014 | CMV-EGFP-IRES-mCherry | Composite | 4 |
| 23 | BBa\_K3338016 | CMV-hGLuc-FLAG | Composite | 4 |
| 24 | BBa\_K3338017 | CMV-hGLuc | Composite | 4 |
| 25 | BBa\_K3338018 | CMV-MagA-FLAG | Composite | 4 |
| 26 | BBa\_K3338019 | IL-6 Pmut-hGLuc | Composite | 4 |
| 27 | BBa\_K3338021 | SynthP\_1-hGLuc | Composite | 4 |
| 28 | BBa\_K3338022 | SynthP\_2-hGLuc | Composite | 4 |
| 29 | BBa\_K3338023 | IL-6 Pmut-MagA-P2A-hGLuc | Plasmid | 4 |
| 30 | BBa\_K3338020 | pEGFP-C2 | Plasmid | 4 |

4.iGEM 부품 사용 사례 분석 4-1. iGEM 페이지 및 할당 프로모터

std\_name <- c("김승화", "김태현","박성군","Aporva","Georgii","유병현","오석진","허성보","이진주")  
igeme.page <- c("https://hayleykim97.github.io/researcheweb/",  
 "https://th-kim310.github.io/researcheweb/",  
 "https://Lelp27.github.io/researcheweb/",  
 "https://aputron.github.io/researcheweb/",  
 "https://gpemelianov.github.io/researcheweb/",  
 "https://yoo-bh.github.io/researcheweb/",  
 "https://seokjin-oh.github.io/researcheweb/",  
 "https://treebird19.github.io/researcheweb/",  
 "https://jinjulee119.github.io/researcheweb/"  
 )  
promoters <- c("BBa\_R0040", "BBa\_R0010", "BBa\_J23100",  
 "BBa\_R0011", "BBa\_I0500", "BBa\_J23101",  
 "BBa\_R0051", "BBa\_J23119", "BBa\_R0062")  
std <- data.frame(std\_name, igeme.page, promoters)  
knitr::kable(std, format = "markdown")

|  |  |  |
| --- | --- | --- |
| std\_name | igeme.page | promoters |
| 김승화 | <https://hayleykim97.github.io/researcheweb/> | BBa\_R0040 |
| 김태현 | <https://th-kim310.github.io/researcheweb/> | BBa\_R0010 |
| 박성군 | <https://Lelp27.github.io/researcheweb/> | BBa\_J23100 |
| Aporva | <https://aputron.github.io/researcheweb/> | BBa\_R0011 |
| Georgii | <https://gpemelianov.github.io/researcheweb/> | BBa\_I0500 |
| 유병현 | <https://yoo-bh.github.io/researcheweb/> | BBa\_J23101 |
| 오석진 | <https://seokjin-oh.github.io/researcheweb/> | BBa\_R0051 |
| 허성보 | <https://treebird19.github.io/researcheweb/> | BBa\_J23119 |
| 이진주 | <https://jinjulee119.github.io/researcheweb/> | BBa\_R0062 |

4-3. create promoter.Rmd file - 해당 팀 정보 - 사용한 부품 및 회로 정보 - 회로 만드는 프로토콜 정보 - 리포터 발현 배양/실험 조건 및 측정방법

1. Analysis of the promoter data

* igem team table
* igem part table
* igem observation table
* table 통합

1. 데이터를 csv에 쓰고 다시 읽기 기존 데이터를 쉽게 excel file로 바꿔준다.
2. 데이터 일관성 유지 (추후 데이터 변환도 함) 테이블 이름 및 데이터 타입 확인

* igem\_team

id (character) team\_name (character) project (character) year (character) wiki (character)

* igem\_part

id (character) BBid (character) type (character) link (character) backbone (character) device\_id (character) (v) team\_name (character) user (character)

* igem\_obs

id (character) strain (character) indc (character) conc (numeric) concunit (character) value (numeric) valunit (character) incubhr (numeric) incubtemps (character) device\_id (character) link (character)

## 연구 방법 2

**웹 페이지 만들기** 1. webpage를 만들기 위해서 Rproject에서 index.Rmd 파일을 만들고 다음과 같이 입력 — title: “My ResearchE Class Website” —

Hello, Website!

## 2.about.Rmd 파일을 생성하고 다음과 같이 입력

## title: “About This Website”

More about this website.

## 3.Result.Rmd 파일을 생성하고 다음과 같이 입력

## title: “Result”

Result

4.\_site.yml 파일을 만들과 다음과 같이 입력 Rstudio > File > New File > Text File 로 만들고 저장을 \_site.yml 로 하기.

name: “My ResearchE Class Website” navbar: title: “My ResearchE Class Website” left: - text: “Home” href: index.html - text: “About” href: about.html

1. Commit in Terminal git add . git commit -m “update” git push
2. 웹 페이지 컴파일 index.Rmd 등 파일들을 knit 버튼으로 html로 컴파일을 하면 필요한 파일들이 들어감
3. html 파일 docs 디렉토리로 복사 \_site 디렉토리 이름을 docs로 바꿔준 수 git tool이나 터미널로 원격저장소에 업데이트 수행

# 연구 결과

1.**기존 download 디렉토리 파일 제거**

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.3 ✓ purrr 0.3.4  
## ✓ tibble 3.1.0 ✓ dplyr 1.0.5  
## ✓ tidyr 1.1.3 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(readxl)  
  
destdir <- "download/"  
filenames <- dir(path = destdir, pattern = "\*.xlsx")  
full\_filenames <- paste0(destdir, "/", filenames)  
sapply(full\_filenames, file.remove)

## download//aputron\_partdb.xlsx download//gpemelianov\_partdb.xlsx   
## TRUE TRUE   
## download//hayleykim97\_partdb.xlsx download//jinjulee119\_partdb.xlsx   
## TRUE TRUE   
## download//Lelp27\_partdb.xlsx download//treebird19\_partdb.xlsx   
## TRUE TRUE

2.**변수 설정**

ids <- c("hayleykim97",   
 "th-kim310",  
 "Lelp27",  
 "aputron",  
 "gpemelianov",  
 "yoo-bh",  
 "seokjin-oh",  
 "treebird19",  
 "jinjulee119"  
 )  
destdir <- "download/"  
  
igem\_team\_cols <- c("id", "team\_name", "project", "year", "wiki")  
igem\_part\_cols <- c("id", "BBid", "type", "link", "backbone", "device\_id", "team\_id", "user")  
igem\_device\_cols <- c("id", "device\_name", "part\_combination")  
igem\_obs\_cols <- c("id", "strain", "indc", "conc", "concunit", "value", "valunit", "incubhr", "incubtemp", "device\_id", "link")

3.**엑셀파일 일괄 다운로드**

for(i in 1:length(ids)){  
 url <- paste0("https://github.com/", ids[i], "/", "researcheweb", "/raw/main/", destdir, "partdb.xlsx")  
 destfile <- paste0(destdir, ids[i], "\_partdb.xlsx")  
 tempfile <- paste0(destdir, "temp\_", ids[i], "\_partdb.xlsx")  
  
   
 ## check the rules   
 ## ===============================================  
 flag <- TRUE  
 try(download.file(url, tempfile, quiet = TRUE), silent=TRUE)  
 #tryCatch(download.file(url, tempfile, quiet = TRUE), error=function(e){})  
 if(!file.exists(tempfile)){  
 print(paste0(i, " ID:", ids[i], " No excel file\n"))  
 flush.console()  
 }else{  
 tmp <- read\_excel(tempfile, sheet = 1, skip = 0, col\_names = T)  
 igem\_cols <- igem\_team\_cols  
 if(length(names(tmp))==length(igem\_cols)){  
 if(!isTRUE(all.equal(names(tmp), igem\_cols))){  
 print(paste(i, ids[i], " check column names in the igem\_team sheet"))  
 flush.console()  
 flag <- FALSE  
 }  
 }else{  
 print(paste(i, ids[i], " check the number of columns in the igem\_team sheet"))  
 flush.console()  
 flag <- FALSE  
 }  
   
 tmp <- read\_excel(tempfile, sheet = 2, skip = 0, col\_names = T)  
 igem\_cols <- igem\_part\_cols  
 if(length(names(tmp))==length(igem\_cols)){  
 if(!isTRUE(all.equal(names(tmp), igem\_cols))){  
 print(paste(i, ids[i], " check column names in the igem\_part sheet"))  
 flush.console()  
 flag <- FALSE  
 }  
 }else{  
 print(paste(i, ids[i], " check the number of columns in the igem\_part sheet"))  
 flush.console()  
 flag <- FALSE  
 }  
   
 tmp <- read\_excel(tempfile, sheet = 3, skip = 0, col\_names = T)  
 igem\_cols <- igem\_device\_cols  
 if(length(names(tmp))==length(igem\_cols)){  
 if(!isTRUE(all.equal(names(tmp), igem\_cols))){  
 print(paste(i, ids[i], " check column names in the igem\_device sheet"))  
 flush.console()  
 flag <- FALSE  
 }  
 }else{  
 print(paste(i, ids[i], " check the number of columns in the igem\_device sheet"))  
 flush.console()  
 flag <- FALSE  
 }  
   
 tmp <- read\_excel(tempfile, sheet = 4, skip = 0, col\_names = T)  
 igem\_cols <- igem\_obs\_cols  
 if(length(names(tmp))==length(igem\_cols)){  
 if(!isTRUE(all.equal(names(tmp), igem\_cols))){  
 cat(i, ids[i], " check column names in the igem\_obs sheet\n")  
 flush.console()  
 flag <- FALSE  
 }  
 }else{  
 print(paste(i, ids[i], " check the number of columns in the igem\_obs sheet"))  
 flush.console()  
 flag <- FALSE  
 }  
   
 if(flag){  
 try(download.file(url, destfile, quiet = TRUE))  
 cat(i, ids[i], " ........... OK\n");flush.console()  
 }  
 file.remove(tempfile)  
 }  
   
   
 ## ===============================================  
 cat("\n");flush.console()  
   
   
}

## 1 hayleykim97 ........... OK

## Warning in download.file(url, tempfile, quiet = TRUE): cannot open URL 'https://  
## github.com/th-kim310/researcheweb/raw/main/download/partdb.xlsx': HTTP status  
## was '404 Not Found'

## [1] "2 ID:th-kim310 No excel file\n"  
##   
## 3 Lelp27 ........... OK  
##   
## 4 aputron ........... OK  
##   
## 5 gpemelianov ........... OK  
##   
## 6 yoo-bh check column names in the igem\_obs sheet  
##   
## 7 seokjin-oh ........... OK  
##   
## 8 treebird19 ........... OK  
##   
## 9 jinjulee119 ........... OK

4.**테이블 통합** 최종 데이터 병합 전 엑셀파일 이름을 사용자 id를 더하여 모든 데이터에 추가 후 데이터 병합

library(magrittr)

##   
## Attaching package: 'magrittr'

## The following object is masked from 'package:purrr':  
##   
## set\_names

## The following object is masked from 'package:tidyr':  
##   
## extract

## 다운로드 받은 엑셀 파일들   
filenames <- dir(path = destdir, pattern = "\*\_partdb.xlsx")  
  
  
tmp1 <- list()  
tmp2 <- list()  
tmp3 <- list()  
tmp4 <- list()  
tmp5 <- list()  
tmp6 <- list()  
tmp7 <- list()  
  
for(i in 1:length(filenames)) {  
 destfile <- paste0(destdir, filenames[i])  
   
 tmp <- read\_excel(destfile, sheet = 1, skip = 0, col\_names = T)  
 tmp %<>% mutate(across(!where(is.character), as.character))   
 ## add filename  
 tmp1[[i]] <- tmp %>% mutate(filename=filenames[i])  
   
 tmp <- read\_excel(destfile, sheet = 2, skip = 0, col\_names = T)  
 tmp %<>% mutate(across(!where(is.character), as.character))   
 tmp2[[i]] <- tmp %>% mutate(filename=filenames[i])  
   
 tmp <- read\_excel(destfile, sheet = 3, skip = 0, col\_names = T)  
 tmp %<>% mutate(across(!where(is.character), as.character))   
 tmp3[[i]] <- tmp %>% mutate(filename=filenames[i])  
   
 tmp <- read\_excel(destfile, sheet = 4, skip = 0, col\_names = T)   
 tmp %<>% mutate(across(!where(is.character), as.character))   
 tmp4[[i]] <- tmp %>% mutate(filename=filenames[i])  
   
   
   
}  
  
igem\_team <- do.call(bind\_rows, tmp1)  
igem\_part <- do.call(bind\_rows, tmp2)  
igem\_device <- do.call(bind\_rows, tmp3)  
igem\_obs <- do.call(bind\_rows, tmp4)

5.**데이터 분석을 위한 전처리** igem\_part와 igem\_team 테이블 병합 id와 file name 변수 두개 동시에 사용

library(tidyverse)  
  
## new id   
tmpdat <- igem\_part %>%   
 left\_join(igem\_team, by=c("team\_id"="id", "filename"="filename"))  
  
tmpdat %>% str

## tibble[,13] [84 × 13] (S3: tbl\_df/tbl/data.frame)  
## $ id : chr [1:84] "1" "2" "3" "4" ...  
## $ BBid : chr [1:84] "BBa\_R0011" "BBa\_C0012" "BBa\_B0054" "BBa\_R0011" ...  
## $ type : chr [1:84] "Promoter" "GFP" "Terminator" "Promoter" ...  
## $ link : chr [1:84] "http://parts.igem.org/Part:BBa\_R0011" "http://parts.igem.org/Part:BBa\_C0012" "http://parts.igem.org/Part:BBa\_B0054" "http://parts.igem.org/Part:BBa\_R0011" ...  
## $ backbone : chr [1:84] "pSB1C3" "pSB1C3" "pSB1C3" "pCDF" ...  
## $ device\_id: chr [1:84] "1" "1" "1" "2" ...  
## $ team\_id : chr [1:84] "1" "1" "1" "2" ...  
## $ user : chr [1:84] "aputron" "aputron" "aputron" "aputron" ...  
## $ filename : chr [1:84] "aputron\_partdb.xlsx" "aputron\_partdb.xlsx" "aputron\_partdb.xlsx" "aputron\_partdb.xlsx" ...  
## $ team\_name: chr [1:84] "ZJU-China" "ZJU-China" "ZJU-China" "UCSF" ...  
## $ project : chr [1:84] "PaDetector, a household device for HPV preliminary screening" "PaDetector, a household device for HPV preliminary screening" "PaDetector, a household device for HPV preliminary screening" "OperationCRISPR" ...  
## $ year : chr [1:84] "2019" "2019" "2019" "2013" ...  
## $ wiki : chr [1:84] "https://2019.igem.org/Team:ZJU-China" "https://2019.igem.org/Team:ZJU-China" "https://2019.igem.org/Team:ZJU-China" "http://2013.igem.org/Team:UCSF" ...

6.**필요한 변수만 선택하는 코드 추가** 필요한 변수는 분석 목적에 따라 다를 수 있음

tmpdat <- igem\_part %>%   
 full\_join(igem\_team, by=c("team\_id"="id", "filename"="filename")) %>%   
 select(id, BBid, type, backbone, device\_id, user, filename, team\_name, year) %>%   
 drop\_na()  
  
tmpdat %>% str

## tibble[,9] [79 × 9] (S3: tbl\_df/tbl/data.frame)  
## $ id : chr [1:79] "1" "2" "3" "4" ...  
## $ BBid : chr [1:79] "BBa\_R0011" "BBa\_C0012" "BBa\_B0054" "BBa\_R0011" ...  
## $ type : chr [1:79] "Promoter" "GFP" "Terminator" "Promoter" ...  
## $ backbone : chr [1:79] "pSB1C3" "pSB1C3" "pSB1C3" "pCDF" ...  
## $ device\_id: chr [1:79] "1" "1" "1" "2" ...  
## $ user : chr [1:79] "aputron" "aputron" "aputron" "aputron" ...  
## $ filename : chr [1:79] "aputron\_partdb.xlsx" "aputron\_partdb.xlsx" "aputron\_partdb.xlsx" "aputron\_partdb.xlsx" ...  
## $ team\_name: chr [1:79] "ZJU-China" "ZJU-China" "ZJU-China" "UCSF" ...  
## $ year : chr [1:79] "2019" "2019" "2019" "2013" ...

7.**관심있는 부품이 사용된 디바이스를 찾고 해당 디바이스 관측 데이터를 수집**

tmpdat2 <- igem\_obs %>%   
 full\_join(igem\_device, by=c("device\_id"="id", "filename"="filename")) %>%   
 drop\_na()  
  
  
tmpdat2 %>% str

## tibble[,14] [88 × 14] (S3: tbl\_df/tbl/data.frame)  
## $ id : chr [1:88] "1" "2" "3" "4" ...  
## $ strain : chr [1:88] "E.coli" "E.coli" "E.coli" "E.coli" ...  
## $ indc : chr [1:88] "Arabinose" "Arabinose" "Arabinose" "Arabinose" ...  
## $ conc : chr [1:88] "0.02" "0.2" "2" "0" ...  
## $ concunit : chr [1:88] "mM" "mM" "mM" "mM" ...  
## $ value : chr [1:88] "3000" "8200" "8000" "250" ...  
## $ valunit : chr [1:88] "Fluorescence" "Fluorescence" "Fluorescence" "a.u." ...  
## $ incubhr : chr [1:88] "12" "12" "12" "4" ...  
## $ incubtemp : chr [1:88] "NA" "NA" "NA" "37" ...  
## $ device\_id : chr [1:88] "1" "1" "1" "2" ...  
## $ link : chr [1:88] "https://2020.igem.org/Team:Jilin\_China/Proof\_Of\_Concept" "https://2020.igem.org/Team:Jilin\_China/Proof\_Of\_Concept" "https://2020.igem.org/Team:Jilin\_China/Proof\_Of\_Concept" "https://2019.igem.org/Team:BHSF\_ND/Project/Leakage#ct3" ...  
## $ filename : chr [1:88] "gpemelianov\_partdb.xlsx" "gpemelianov\_partdb.xlsx" "gpemelianov\_partdb.xlsx" "gpemelianov\_partdb.xlsx" ...  
## $ device\_name : chr [1:88] "D0001" "D0001" "D0001" "D0002" ...  
## $ part\_combination: chr [1:88] "1,2,3,4" "1,2,3,4" "1,2,3,4" "5,6" ...

8.**관심있는 part가 사용된 device와 관측값을 찾기 위한 데이터 처리 수행**

tmpdat %>%   
 filter(BBid=="BBa\_R0040")

## # A tibble: 5 x 9  
## id BBid type backbone device\_id user filename team\_name year   
## <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 5 BBa\_R0… promot… PSB3C5 1 hayley… hayleykim97\_… Fudan 2019   
## 2 14 BBa\_R0… promot… PSB3C5 2 hayley… hayleykim97\_… Fudan 2019   
## 3 23 BBa\_R0… promot… PSB3C5 3 hayley… hayleykim97\_… Fudan 2019   
## 4 28 BBa\_R0… promot… PSB3C5 4 hayley… hayleykim97\_… Fudan 2019   
## 5 10 BBa\_R0… promot… PSB3C5 3 JinjuL… jinjulee119\_… Fudan 2019

수집 1

finaldat <-tmpdat2 %>%   
 mutate(partcomb = lapply(strsplit(tmpdat2$part\_combination, split=","), as.numeric)) %>%   
 filter(unlist(lapply(partcomb, function(x){5 %in% x})) & filename=="hayleykim97\_partdb.xlsx")  
  
finaldat %>% str

## tibble[,15] [14 × 15] (S3: tbl\_df/tbl/data.frame)  
## $ id : chr [1:14] "1" "2" "3" "4" ...  
## $ strain : chr [1:14] "E.coli" "E.coli" "E.coli" "E.coli" ...  
## $ indc : chr [1:14] "tetR" "tetR" "tetR" "tetR" ...  
## $ conc : chr [1:14] "300" "450" "600" "750" ...  
## $ concunit : chr [1:14] "ng/ml" "ng/ml" "ng/ml" "ng/ml" ...  
## $ value : chr [1:14] "10000" "110000" "175000" "160000" ...  
## $ valunit : chr [1:14] "MEFL/ particle" "MEFL/ particle" "MEFL/ particle" "MEFL/ particle" ...  
## $ incubhr : chr [1:14] "7" "7" "7" "7" ...  
## $ incubtemp : chr [1:14] "37" "37" "37" "37" ...  
## $ device\_id : chr [1:14] "1" "1" "1" "1" ...  
## $ link : chr [1:14] "Part:BBa K3245007 - parts.igem.org" "Part:BBa K3245007 - parts.igem.org" "Part:BBa K3245007 - parts.igem.org" "Part:BBa K3245007 - parts.igem.org" ...  
## $ filename : chr [1:14] "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" ...  
## $ device\_name : chr [1:14] "BBa\_K3245007" "BBa\_K3245007" "BBa\_K3245007" "BBa\_K3245007" ...  
## $ part\_combination: chr [1:14] "1,2,3,4,5,6,7,8,9" "1,2,3,4,5,6,7,8,9" "1,2,3,4,5,6,7,8,9" "1,2,3,4,5,6,7,8,9" ...  
## $ partcomb :List of 14  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9  
## ..$ : num [1:9] 1 2 3 4 5 6 7 8 9

수집 2

finaldat <-tmpdat2 %>%   
 mutate(partcomb = lapply(strsplit(tmpdat2$part\_combination, split=","), as.numeric)) %>%   
 filter(unlist(lapply(partcomb, function(x){14 %in% x})) & filename=="hayleykim97\_partdb.xlsx")  
  
finaldat %>% str

## tibble[,15] [10 × 15] (S3: tbl\_df/tbl/data.frame)  
## $ id : chr [1:10] "15" "16" "17" "18" ...  
## $ strain : chr [1:10] "E.coli" "E.coli" "E.coli" "E.coli" ...  
## $ indc : chr [1:10] "tetR" "tetR" "tetR" "tetR" ...  
## $ conc : chr [1:10] "150" "300" "450" "600" ...  
## $ concunit : chr [1:10] "ng/ml" "ng/ml" "ng/ml" "ng/ml" ...  
## $ value : chr [1:10] "125000" "2e+05" "210000" "250000" ...  
## $ valunit : chr [1:10] "MEFL/ particle" "MEFL/ particle" "MEFL/ particle" "MEFL/ particle" ...  
## $ incubhr : chr [1:10] "7" "7" "7" "7" ...  
## $ incubtemp : chr [1:10] "37" "37" "37" "37" ...  
## $ device\_id : chr [1:10] "2" "2" "2" "2" ...  
## $ link : chr [1:10] "Part:BBa K3245012 - parts.igem.org" "Part:BBa K3245012 - parts.igem.org" "Part:BBa K3245012 - parts.igem.org" "Part:BBa K3245012 - parts.igem.org" ...  
## $ filename : chr [1:10] "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" ...  
## $ device\_name : chr [1:10] "BBa\_K3245012" "BBa\_K3245012" "BBa\_K3245012" "BBa\_K3245012" ...  
## $ part\_combination: chr [1:10] "11,12,13,14,15,16,17,18" "11,12,13,14,15,16,17,18" "11,12,13,14,15,16,17,18" "11,12,13,14,15,16,17,18" ...  
## $ partcomb :List of 10  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18  
## ..$ : num [1:8] 11 12 13 14 15 16 17 18

수집 3

finaldat <-tmpdat2 %>%   
 mutate(partcomb = lapply(strsplit(tmpdat2$part\_combination, split=","), as.numeric)) %>%   
 filter(unlist(lapply(partcomb, function(x){23 %in% x})) & filename=="hayleykim97\_partdb.xlsx")  
  
finaldat %>% str

## tibble[,15] [10 × 15] (S3: tbl\_df/tbl/data.frame)  
## $ id : chr [1:10] "25" "26" "27" "28" ...  
## $ strain : chr [1:10] "E.coli" "E.coli" "E.coli" "E.coli" ...  
## $ indc : chr [1:10] "tetR" "tetR" "tetR" "tetR" ...  
## $ conc : chr [1:10] "150" "300" "450" "600" ...  
## $ concunit : chr [1:10] "ng/ml" "ng/ml" "ng/ml" "ng/ml" ...  
## $ value : chr [1:10] "4e+05" "350000" "295000" "3e+05" ...  
## $ valunit : chr [1:10] "MEFL/ particle" "MEFL/ particle" "MEFL/ particle" "MEFL/ particle" ...  
## $ incubhr : chr [1:10] "7" "7" "7" "7" ...  
## $ incubtemp : chr [1:10] "37" "37" "37" "37" ...  
## $ device\_id : chr [1:10] "3" "3" "3" "3" ...  
## $ link : chr [1:10] "Part:BBa K3245011 - parts.igem.org" "Part:BBa K3245011 - parts.igem.org" "Part:BBa K3245011 - parts.igem.org" "Part:BBa K3245011 - parts.igem.org" ...  
## $ filename : chr [1:10] "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" ...  
## $ device\_name : chr [1:10] "BBa\_K3245011" "BBa\_K3245011" "BBa\_K3245011" "BBa\_K3245011" ...  
## $ part\_combination: chr [1:10] "19,20,21,22,23,24,25,26,27" "19,20,21,22,23,24,25,26,27" "19,20,21,22,23,24,25,26,27" "19,20,21,22,23,24,25,26,27" ...  
## $ partcomb :List of 10  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27  
## ..$ : num [1:9] 19 20 21 22 23 24 25 26 27

수집 4

finaldat <-tmpdat2 %>%   
 mutate(partcomb = lapply(strsplit(tmpdat2$part\_combination, split=","), as.numeric)) %>%   
 filter(unlist(lapply(partcomb, function(x){28 %in% x})) & filename=="hayleykim97\_partdb.xlsx")  
  
finaldat %>% str

## tibble[,15] [10 × 15] (S3: tbl\_df/tbl/data.frame)  
## $ id : chr [1:10] "35" "36" "37" "38" ...  
## $ strain : chr [1:10] "E.coli" "E.coli" "E.coli" "E.coli" ...  
## $ indc : chr [1:10] "tetR" "tetR" "tetR" "tetR" ...  
## $ conc : chr [1:10] "150" "300" "450" "600" ...  
## $ concunit : chr [1:10] "ng/ul" "ng/ul" "ng/ul" "ng/ul" ...  
## $ value : chr [1:10] "370000" "370000" "370000" "390000" ...  
## $ valunit : chr [1:10] "MEFL/ particle" "MEFL/ particle" "MEFL/ particle" "MEFL/ particle" ...  
## $ incubhr : chr [1:10] "5" "5" "5" "5" ...  
## $ incubtemp : chr [1:10] "37" "37" "37" "37" ...  
## $ device\_id : chr [1:10] "4" "4" "4" "4" ...  
## $ link : chr [1:10] "Part:BBa K3245013 - parts.igem.org" "Part:BBa K3245013 - parts.igem.org" "Part:BBa K3245013 - parts.igem.org" "Part:BBa K3245013 - parts.igem.org" ...  
## $ filename : chr [1:10] "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" "hayleykim97\_partdb.xlsx" ...  
## $ device\_name : chr [1:10] "BBa\_K3245013" "BBa\_K3245013" "BBa\_K3245013" "BBa\_K3245013" ...  
## $ part\_combination: chr [1:10] "28,29,30,31" "28,29,30,31" "28,29,30,31" "28,29,30,31" ...  
## $ partcomb :List of 10  
## ..$ : num [1:4] 28 29 30 31  
## ..$ : num [1:4] 28 29 30 31  
## ..$ : num [1:4] 28 29 30 31  
## ..$ : num [1:4] 28 29 30 31  
## ..$ : num [1:4] 28 29 30 31  
## ..$ : num [1:4] 28 29 30 31  
## ..$ : num [1:4] 28 29 30 31  
## ..$ : num [1:4] 28 29 30 31  
## ..$ : num [1:4] 28 29 30 31  
## ..$ : num [1:4] 28 29 30 31

수집 5

finaldat <-tmpdat2 %>%   
 mutate(partcomb = lapply(strsplit(tmpdat2$part\_combination, split=","), as.numeric)) %>%   
 filter(unlist(lapply(partcomb, function(x){10 %in% x})) & filename=="jinjulee119\_partdb.xlsx")  
  
finaldat %>% str

## tibble[,15] [4 × 15] (S3: tbl\_df/tbl/data.frame)  
## $ id : chr [1:4] "1" "2" "3" "4"  
## $ strain : chr [1:4] "E. coli Nissle 1917" "E. coli Nissle 1917" "E. coli Nissle 1917" "E. coli Nissle 1917"  
## $ indc : chr [1:4] "tet" "tet" "tet" "tet"  
## $ conc : chr [1:4] "600" "750" "900" "1200"  
## $ concunit : chr [1:4] "ng/㎕" "ng/㎕" "ng/㎕" "ng/㎕"  
## $ value : chr [1:4] "350000" "230000" "180000" "150000"  
## $ valunit : chr [1:4] "MEFL/particle" "MEFL/particle" "MEFL/particle" "MEFL/particle"  
## $ incubhr : chr [1:4] "7" "7" "7" "7"  
## $ incubtemp : chr [1:4] "37" "37" "37" "37"  
## $ device\_id : chr [1:4] "3" "3" "3" "3"  
## $ link : chr [1:4] "https://2019.igem.org/Team:Fudan/" "https://2019.igem.org/Team:Fudan/" "https://2019.igem.org/Team:Fudan/" "https://2019.igem.org/Team:Fudan/"  
## $ filename : chr [1:4] "jinjulee119\_partdb.xlsx" "jinjulee119\_partdb.xlsx" "jinjulee119\_partdb.xlsx" "jinjulee119\_partdb.xlsx"  
## $ device\_name : chr [1:4] "D0003" "D0003" "D0003" "D0003"  
## $ part\_combination: chr [1:4] "10,11,12,13" "10,11,12,13" "10,11,12,13" "10,11,12,13"  
## $ partcomb :List of 4  
## ..$ : num [1:4] 10 11 12 13  
## ..$ : num [1:4] 10 11 12 13  
## ..$ : num [1:4] 10 11 12 13  
## ..$ : num [1:4] 10 11 12 13

9.**그래프 그리는데 필요한 변수 추출 후 정리**

plotdat <- finaldat %>%   
 select(-c(id, link, filename, part\_combination, partcomb)) %>%   
 mutate(value = as.numeric(value))  
  
plotdat %>% str

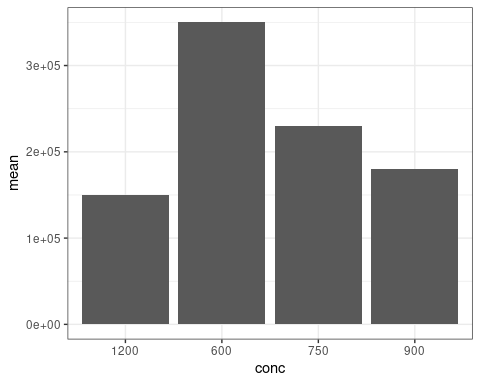
## tibble[,10] [4 × 10] (S3: tbl\_df/tbl/data.frame)  
## $ strain : chr [1:4] "E. coli Nissle 1917" "E. coli Nissle 1917" "E. coli Nissle 1917" "E. coli Nissle 1917"  
## $ indc : chr [1:4] "tet" "tet" "tet" "tet"  
## $ conc : chr [1:4] "600" "750" "900" "1200"  
## $ concunit : chr [1:4] "ng/㎕" "ng/㎕" "ng/㎕" "ng/㎕"  
## $ value : num [1:4] 350000 230000 180000 150000  
## $ valunit : chr [1:4] "MEFL/particle" "MEFL/particle" "MEFL/particle" "MEFL/particle"  
## $ incubhr : chr [1:4] "7" "7" "7" "7"  
## $ incubtemp : chr [1:4] "37" "37" "37" "37"  
## $ device\_id : chr [1:4] "3" "3" "3" "3"  
## $ device\_name: chr [1:4] "D0003" "D0003" "D0003" "D0003"

10.**bar graph를 그리는 코드 추가** 균주, inducer, inducer 농도별로 평균을 계산 후 그래프 작성

datasummary <- plotdat %>%   
 group\_by(indc, conc) %>%   
 summarise(mean=mean(value), n=n())

## `summarise()` has grouped output by 'indc'. You can override using the `.groups` argument.

ggplot(datasummary, aes(x=conc, y=mean)) +  
 geom\_bar(stat="identity") +  
 theme\_bw()



# 고찰

전혀 다룰 줄 몰랐던 R tool을 처음 배우며 웹페이지 만드는 법, 코드 청크 이해(?) 등 tool을 다루면서 기존 방식과는 다른 방식으로 데이터를 수집하고 다시 해석할 수 있는 기반 지식을 알게 되었다. R tool에 대해서는 기초지식조차 없어서 수업 진도를 따라가는데 많이 뒤쳐지기는 했었지만 충분히 해결할 수 있는 문제들이었다.