Initial Project Ideas

Salome Steinke

26.04.2019

# Initial Project Ideas

Type of cancer chosen: Breast Cancer

## The Dataset

Loading Dataset as follows:

allDepMapData = readRDS("~/4.FS/Bioinfo/DepMap19Q1\_allData.RDS")  
names(allDepMapData)

## [1] "expression" "copynumber" "mutation" "kd.ceres" "kd.prob"   
## [6] "annotation"

The list named allDepMapData contains the following matrices:

* **Gene expression**: consists of TPM (transcripts per million) values reflecting level of gene expression. Higher values suggest overexpression of genes. Matrix rows are gene names and columns cancer cell line identifiers. Column and row names *character*; TPM values *integer*

dim(allDepMapData$expression)

## [1] 49070 544

typeof(col(allDepMapData$expression))

## [1] "integer"

* **Gene copy number**: consists of gene copy number (CN) values. When two alleles per gene present CN = 2, deletions shown by CN < 2 and amplifications CN > 2. Matrix rows are gene names and columns cancer cell line identifiers.
* **Gene mutations**: Annotation file for various mutations observed in sample. isDeleterious flag specificies if mutation has functional effect or not.
* **Gene knockdown (CERES)**: Gene knockdown scores. These scores measure how essential a particular gene is to cell survival, by reflecting whether a knockdown of the gene reduces , increases or leads to no change of cell proliferation. Smaller values refer to higher importance. Matrix rows are gene names and columns cancer cell line identifiers.

dim(allDepMapData$kd.ceres)

## [1] 17634 544

* **Gene knockdown (Probability)**: The probabilty for the effect of gene knockdown, with higher probabilities signifying a high likelyhood of the gene reducing cell proliferation.

dim(allDepMapData$kd.prob)

## [1] 17634 544

* **Annotation**: This matrix gives information regarding cell lines. The CCLE\_Name encoded as cell line name \_ Tissue of origin. The DepMap\_IDgives uniform cell line identifiers used in all data sets. The columns Primary.Disease and Subtype.Disease refer to tissue/tumor of origin.

dim(allDepMapData$annotation)

## [1] 544 7

colnames(allDepMapData$annotation)

## [1] "DepMap\_ID" "CCLE\_Name" "Aliases" "Primary.Disease"  
## [5] "Subtype.Disease" "Gender" "Source"

rownames(allDepMapData$annotation)

## [1] "4" "5" "7" "8" "10" "11" "12" "13" "14" "16"   
## [11] "17" "18" "20" "21" "24" "27" "28" "29" "34" "35"   
## [21] "36" "38" "39" "40" "41" "44" "46" "52" "53" "54"   
## [31] "59" "66" "73" "76" "78" "80" "83" "84" "89" "92"   
## [41] "94" "95" "96" "101" "102" "105" "110" "113" "115" "117"   
## [51] "120" "124" "125" "129" "134" "135" "139" "140" "143" "144"   
## [61] "146" "147" "148" "149" "152" "155" "156" "163" "164" "165"   
## [71] "174" "179" "184" "185" "186" "193" "194" "199" "200" "202"   
## [81] "203" "204" "207" "211" "216" "217" "218" "221" "222" "223"   
## [91] "226" "227" "228" "229" "230" "232" "233" "237" "238" "239"   
## [101] "241" "243" "245" "246" "247" "248" "250" "251" "252" "253"   
## [111] "254" "255" "256" "257" "258" "259" "260" "261" "267" "269"   
## [121] "272" "273" "276" "277" "283" "286" "288" "289" "290" "291"   
## [131] "297" "299" "302" "303" "305" "306" "307" "310" "311" "312"   
## [141] "313" "314" "315" "317" "318" "319" "324" "325" "326" "327"   
## [151] "329" "330" "331" "332" "334" "336" "343" "345" "347" "351"   
## [161] "355" "356" "357" "358" "359" "360" "362" "367" "368" "369"   
## [171] "374" "376" "380" "382" "383" "386" "388" "389" "390" "393"   
## [181] "394" "395" "396" "398" "399" "400" "402" "403" "406" "407"   
## [191] "408" "412" "413" "418" "419" "420" "426" "428" "429" "437"   
## [201] "438" "441" "443" "445" "447" "451" "452" "453" "456" "457"   
## [211] "459" "460" "461" "462" "463" "464" "465" "466" "467" "468"   
## [221] "469" "471" "472" "473" "475" "476" "477" "478" "479" "480"   
## [231] "482" "483" "486" "487" "490" "491" "493" "494" "496" "498"   
## [241] "501" "502" "503" "507" "508" "509" "511" "512" "513" "515"   
## [251] "518" "519" "528" "532" "533" "535" "536" "538" "539" "540"   
## [261] "541" "543" "548" "560" "561" "562" "564" "567" "570" "571"   
## [271] "575" "576" "579" "580" "582" "584" "591" "592" "596" "597"   
## [281] "599" "605" "607" "609" "610" "611" "612" "613" "614" "615"   
## [291] "618" "621" "622" "623" "624" "628" "629" "637" "638" "639"   
## [301] "640" "642" "643" "644" "648" "652" "653" "654" "656" "662"   
## [311] "663" "664" "667" "668" "672" "674" "675" "679" "684" "685"   
## [321] "686" "687" "688" "690" "691" "692" "697" "702" "704" "705"   
## [331] "709" "711" "712" "715" "716" "723" "725" "726" "727" "729"   
## [341] "737" "738" "739" "744" "747" "748" "749" "750" "752" "753"   
## [351] "754" "755" "756" "757" "758" "759" "760" "761" "763" "766"   
## [361] "767" "768" "771" "773" "774" "775" "777" "778" "781" "782"   
## [371] "784" "786" "788" "789" "790" "792" "794" "795" "799" "800"   
## [381] "801" "807" "808" "809" "810" "811" "812" "813" "816" "817"   
## [391] "819" "820" "821" "822" "823" "824" "825" "826" "827" "828"   
## [401] "829" "831" "834" "835" "836" "837" "839" "843" "844" "846"   
## [411] "847" "849" "850" "853" "854" "855" "857" "860" "861" "862"   
## [421] "864" "866" "868" "869" "870" "872" "873" "875" "877" "884"   
## [431] "886" "888" "889" "893" "894" "895" "898" "899" "900" "901"   
## [441] "902" "903" "904" "907" "908" "910" "912" "914" "915" "918"   
## [451] "921" "922" "923" "925" "927" "929" "932" "933" "934" "936"   
## [461] "939" "940" "943" "944" "949" "954" "955" "958" "959" "960"   
## [471] "961" "964" "970" "972" "974" "979" "980" "982" "994" "996"   
## [481] "999" "1000" "1001" "1003" "1012" "1014" "1016" "1017" "1026" "1030"  
## [491] "1043" "1053" "1056" "1064" "1065" "1070" "1084" "1085" "1088" "1091"  
## [501] "1092" "1097" "1098" "1099" "1101" "1102" "1108" "1113" "1114" "1116"  
## [511] "1118" "1121" "1123" "1125" "1126" "1127" "1128" "1130" "1131" "1135"  
## [521] "1141" "1144" "1147" "1157" "1159" "1160" "1171" "1172" "1175" "1178"  
## [531] "1181" "1190" "1205" "1220" "1221" "1226" "1238" "1240" "1246" "1253"  
## [541] "1262" "1264" "1304" "1337"

## Questions so far

1. Which genes are in dataset?
2. Which genes are we interested in?
3. What are cell line identifiers? Which ones do we have? Perhaps analyse using the *Annotation Matrix*
4. Analyse TPM levels for selected genes and cell line identifiers *(Gene expression matrix)*
5. Analyse *Gene knockdown matrix* to check for importance of genes in cancer cell lines for survival. Compare to *Gene knockdown probability* to check for probability of reducing/increasing cell proliferation.
6. Analyse gene copy numbers for genes and cell line identifiers to identify in which cell lines mutations have occured (compare gene copy numbers for different cell lines) *(Gene copy number matrix)*
7. In cell lines where mutations detected check if they have functional effect or not *(Gene mutations matrix)*
8. which are importat driver mutations in breast cancer? Which are the most important passenger mutations associated with them?
9. ist there an association between driver and passenger alterations? Could synthetic lethal interactions be established by interfering in the interactions between driver and passenger genes? 10.are there any drugs availabe so far targeting the mutations?
10. how could the obtained data be used for cancer therapy?

## Things to do

* identify 3-5 most prominent mutations/genetic alterations (ideally those difficult to target clinically)
* identify appropriate cell line models
* read Papers and extract interesting questions, methods, analysis
* Questions with milestones/ timeframe
* Data clean up

#finding all breast cancer cell lines  
which(allDepMapData$annotation[4]== "Breast Cancer")

## [1] 10 12 16 41 60 87 102 110 121 125 148 159 168 267 287 290 324  
## [18] 328 333 335 355 364 406 409 411 421 437 532

Breast\_Cancer\_Celllines <- subset(allDepMapData$annotation, Primary.Disease == "Breast Cancer")  
#see what we got  
summary(Breast\_Cancer\_Celllines)

## DepMap\_ID CCLE\_Name Aliases   
## Length:28 Length:28 Length:28   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Primary.Disease  
## Breast Cancer :28   
## Bladder Cancer : 0   
## Bone Cancer : 0   
## Brain Cancer : 0   
## Colon/Colorectal Cancer : 0   
## Endometrial/Uterine Cancer: 0   
## (Other) : 0   
## Subtype.Disease Gender   
## Carcinoma :14 : 0   
## Breast Ductal Carcinoma :11 -1 : 0   
## Adenocarcinoma : 1 Female:28   
## Breast Ductal Carcinoma, medullary : 1 Male : 0   
## Carcinoma, anaplastic : 1   
## Acute Lymphoblastic Leukemia (ALL), B-cell: 0   
## (Other) : 0   
## Source   
## ATCC :22   
## DSMZ : 4   
## : 1   
## HSSRB : 1   
## Academic Lab: 0   
## EACC : 0   
## (Other) : 0

* Data exploration (Descriptive statistics, graphical representations of data and interdependancies )
* Dimensionality reduction (PCA? ; kmeans or clustering of samples)
* find which are most correlated mutations/copy number changes
* functional/biological relationship between correlated alterations
* find clinically approved drugs for correlated alterations
* Data modelling (prediction of cell line sensitivity to knockdown or impact of gene mutation/copy number on gene expression)

# NOTES ON RESEARCH PAPERS

### BREAST CANCER MUTATIONS

### A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers (Berger et al., 2018):

* Focus on invasive breast carcinoma (BRCA) (1087 Patients)
* Mutations: TP53; PIK3CA; MYC, CCND1
* ARID1A, ERBB3, BRCA1, FBXW7, KMT2C, PIK3CA, PIK3R1, PPP2R1A, PTEN, and TP53 were mutated at higher frequencies across the Pan-Gyn tumor types than across the non-Gyn types (false discovery rate [FDR] < 0.01, Fisher’s exact test) (Figure 1A)
* The top ﬁve most frequently mutated genes were TP53 (44% of samples mutated), PIK3CA (32%), PTEN (20%), ARID1A (14%), and PIK3R1 (11%)
* Correlation coefficient calculations with heat maps and tree diagrams (unsupervised hierarchical clustering) showing clusters
* Mutation signatures- unique mutation types – displayed using six substitution subtypes
  + <https://cancer.sanger.ac.uk/cosmic/signatures>
  + Breast cancer mutation signatures: Signature 1, signature 3, signature 5 (aetiology unknown), signature 8, signature 17 and signature 18 (aetiology unknown), signature 20, signature 26
  + May be interesting to compare identified mutations in genes from own dataset to the substitution types known for breast cancer
* Breast cancer associated with COSMIC 5
* Methods used:
  + Fisher’s Extact test with Bejamini-Hochberg method for adjusted for false discovery rates
  + Mann Whitney U test to identify genes with significant differs between meidain methylation levels of genes between two groups
  + MutSigCV v1.4 <www.broadinstitute.org>
  + Silhouette analysis
  + T-test and Wilcoxon Rank Sum Test
  + Survival analysis with R package “survival”
    - Kaplan-Meier or Cox Proportional Hazards Models
    - For more info: <https://www.datacamp.com/community/tutorials/survival-analysis-R>

### Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer (Ciriello et al., 2015):

* 817 breast tumor samples
* Invasive lobular carcinoma (ILC)- second most prevalent histologic subtype of invasive breast cancer (10-15% oof all cases)
  + Most were identified as Luminal
* E-cadherin (CDH1) loss in 90% of cases
  + Rare HER2 protein overexpression or amplification
  + Tumor invasion and metastasis in other tumor types including diffuse gastric cancer
* ILC, IDC, mixed and other carcinoma types
* Mutations targeting PTEN, TBX3, FOXA1, PIK3CA, CDH1
  + PTEN loss associated with increased AKT phosphorylation
  + GATA3 mutations- high expression luminal A IDC- differential modulation of ER activity in ILC/IDC
  + TP53 und PIK3CA in ILC und IDC as well as MYC
  + ILC cases enriched in CDH1 mutations
  + Recurrently mutated genes identified by MutSigCV2
  + Recurrent copy-number alterations estimated by GISTIC
* FOXA1
  + Key ER transcriptional modulator coordinating ER DAN binding within a large protein cmplex by modifying chromatin accessibility and mediating long-range DNA interactions
  + Mutation cluster in loop 2 of forkhead-DNA (three a-helices, 3 ß-strands and two loops) binding and C-term. transactivation domains
* PTEN
  + Homozygous deletions, somatic mutations and mutually exclusive with PIK3CA mutations
  + Lower PTEN protein in LumA ILC than LumA IDC
  + PTEN protein as negative regulator of Akt activity
* 68 mutated genes, 47 regions of gain, 63 regions of loss identified
* Genetic difference between ER+/luminal and ER-/basal-like breast cancer
* HER2+ breast tumours

### Comprehensive molecular portraits of human breast tumours (The Cancer Genome Atlas et al., 2012) <https://doi.org/10.1038/nature11412>:

* TP53, PIK3CA and GATA3 with subtype-associated mutations in GATA3, PIK3CA and MAP3K1 with luminal A subtype
* Similar aetiology and therapeutic opportunities with ovarian cancer
* Three groups of breast cancer
  + Oestrogen receptor (ER) positive group most numerous and diverse
  + HER2 (or ERBB2) amplified group- great clinical success in targeting HER2
  + Basal-like breast cancers (triple-negative breast cancers) lack expression of ER, progesterone receptor (PR) and HER2
  + PIK3CA, PTEN, AKT1, TP53, GATA3, CDH1, RB1, MLL3, MAP3K1 and CDKN1B
  + With novel mutations: TBX3, RUNX1, CBFB, AFF2, PIK3R1, PTPN22, PTPRD, NF1, SF3B1 and CCND3
  + FOXA1 and CTCF transcription factor mutations
  + The luminal A subtype harboured the most significantly mutated genes, with the most frequent being PIK3CA (45%), followed by MAP3K1, GATA3, TP53, CDH1 and MAP2K4. (missense)
  + Luminal B cancers : TP53 and PIK3CA (29% each) most frequent (missense)
  + Basal like cancers: TP53 (nonsense, frame shift)
  + HER2E subtype: high frequency of TP53 (72%) and PIK3CA (39%) mutations + lower frequency with PIK3R1 (4%)
* Luminal/ER+:
  + ESR1, GATA3, FOXA1, XBP1 and MYB
  + GATA3 and FOXA1 mutated in mutually exclusive fashion
  + PIK3CA high mutation frequency
  + Frequent MAP3K1 and MAP2K4 mutation
  + MEMo analysis: all MAP3K1 and MAP2K4 mutations were in luminal tumors (almost mutually exclusive to each other though
  + TP53 highest in luminal A cancers
  + Hyperactivation transcriptional activity with MYC and FOXMI proliferation
  + Luminal A tumours- retain RB1 and TP53 activity
* BRCA1/BRCA2
* Basel-like cancers
* TP53, RB1 and BRCA1 mutations
* notes from elias, just to not erase them :)
* Methods used were genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays
* The analysis revealed 4 major breast cancer classes, which are different in gene expression
* In therapy 3 major classes are distinguished: ER+ Patients, which receive endocrine therapy, HER2 amplification patients (we had strategies treatment strategies in lecture) and triple negative breast cancer (TNBC) patients, often BRCA1 is mutated, patients can only be treated with chemotherapy
* The new step of this paper is to include the methods mentioned above. Earlier studies just used data from mRNA expression, DNA copies and parallel sequencing (no idea what the last thing is)
* Total number of patients is 825
* A LOT of different mutations were found during whole exome sequencing, which lead to following new genes associated with breast cancer: TBX3, RUNX1, CBFB, AFF2, PIK3R1, PTPN22, PTPRD, NF1, SF3B1 and CCND3.
* Genes confirmed and already known to be important for breast cancer development were: PIK3CA, PTEN, AKT1, TP53, GATA3, CDH1, RB1, MLL3, MAP3K1 and CDKN1B.
* mRNA-expression of cancer subtypes revealed different mutation patterns in different subclasses ( needs a more detailed description)
* Patterns vary also by mutation type (nonsense, misssense etc.)
* Further analysis showed a strong connection to the germ line of ~10% of breast cancers, including BRCA1 mutations, confirming the TNBC subclass (not the best description I ever made 😊)

###Genetic Interactions in Cancer Progression and Treatment (Ashworth et al.2011) DOI: 10.1016/j.cell.2011.03.020

### Predicting Cancer-Speciﬁc Vulnerability via Data-Driven Detection of Synthetic Lethality:

* Synthetic lethality: inhibition of two nonessential genes lethal while inhibition of a single gene is not
* Targeting synthetic lethal (SL) partners in treatment
* Use of DAISY (data mining synthetic lethality identification pipeline) ++ Genomic survival of the fittest (SoF) based on observation that cancer cells having lost two SL-paried genes don’t survive – SL cells elimination form population hence somatic copy number alterations (SCNA) and somatic mutation data can detect this inactivation events ++ SL pairs of given gene can be detected underexpression or low copy number induces essentiality of the gene ++ Pairwise gene coexpression ++ Gene pairs fulfilling all three of the above ++ Detection of overactivity also possible by selecting two genes if expression is correlated and if overactivity of one induces essentiality of the other (first examining if gene B becomes essential when gene A is overactive; second if gene B has a higher SCNA level when gene A is overactive) ++ Returns p values denoting significance of SL or SDL interaction
* Wilcoxon rank sum p values
* Predicting gene essentiality is cell-line-specific
* What are ROC curves?
* Signed Kaplan-Meier (KM) score
* Experimental procedures section most interesting for us: ++ (1) Wilcoxon rank sum test to determine of gene B significantly higher SCNA levels when gene A is inactive/overactive in rest of samples. Ouput= gene pairs that pass the test with specific significant level ++ (2) input: gene essentiality data with SCNA and gene expression profiles. Wilcoxon rank sum test to determine if Gene B lower scores where gene A inactive/overactive ++ P value < 0.05 used with Bonferroni correction ++ Gene defined underexpressed if SCAN below -0.3; likewise overactive when above 0.3 ++ Generally genes in sample underexpressed if in 10th percentile of all expression levels and overexpressed if above the 90the percentile ++ For clustered genes and gene maps visit < <http://www.cs.tau.ac.il/livnatje/SL_network.zip> >

### Review: Synthetic lethality and cancer (O’Neill, 2017):

* **synthetic lethal interaction (SLI)**: interaction of two genes which is only lethal if both genes are perturbated BUT NOT if only one gene is perturbated  
  ++ involves cancer-specific interaction which is “non-oncogene addiction” + one tumour-specific somatic mutation ++ mutations may contribute both to tumour proliferation and/or an increase in tumour vulnerability to tgerapeutics ++ therapeutical concept: targeting of the passenger mutation (= second-site synthetic lethal target) & overcoming undruggable “oncogene addiction” mutations
* **challenges in identfying SLI**: ++ SLI may be condition-dependend (both refering to genetic background & cellular conditions) ++ SLI are rare ++ identification of mutants difficult as cells die due to SLI which impedes mutant recovery
* **types of synthetic lethality**: ++ synthetic dosage lethality (SDL): overexpression of one gene (due to epigenetic gain-of-function mutation or increase in gene copy number) interacting with another gene reduced in function results in lethality (breast cancer: SDL of high CKS1B with inhibition/knockdown of PLK1) ++ conditional synthetic lethality: intrinisic (genetic) and extrinsinc (tumour microenvironment, i.e. hypoxia) may be predispositions for efficiency of SLI, thus alteration of tumour-specific conditions may be useful in order to increase cytotoxicity of therapeutics (breast cancer: use of PARP-inhibitors in homologous recombination deficient tumours, also SLI of PARP with BRCA1)
* **screening for SLI in model organisms**: ++ use of high-throughput approaches in yeast based on existing model organis interaction networks for SLI with cancer-relevant homologues, if no homologues available then screening for cancer-relevant pathways & disruption of gene involved in a similar process in humans ++ use of synthetic lethality genetic networks in yeast: network based on quantification of digenic interactions generated by double mutant construction -> identification of > 500.00 SLIs, findings may be used for identifiation of human (homologue & non-homologue) SLI candidates due to evolutionary conservation  
  ++ SLI screening in human cells: CRISPR outperforms RNAi due to greater RNA libraries & stability of CRISPR based gene knockouts, BUT cell line specificty & targeting (RNAi: due to interference with endogenous RNAi pathways, CRISPR: due to off-target effects & toxicity in aneuploid cells) remain problematic -> use of CRISPR gene editing, CRISPRi & CRISPRa (fusion with repressor/activator), CombiGEM-CRISPR (multiplexed sgRNAs targeting up to 6 sequences), Pertub-seq (CRISPR + single-cell transriptome analysis instead of effects on cellular fitness) ++ predicting SLI candidates based on statistical genetics: search for mutually exclusive (= SLI suitable) mutations in databases (e.g. The Cancer Genome Atlas, Calatolgue of Somatic Mutations in Cancer) & use of DAISY (bioinformatic approach using genes essentiality profiles, expression data & data on copy number alterations)
* **use of SLI for cancer therapy**: ++ SLI reported in SLI databases (e.g. SynLethDB) ++ therapeutic approaches may involve increasing the tumour susceptibility to conventional treatments (i.e. chemotherapy) & exploiting SLI by manipulation of the second-side targets (passenger mutation, e.g. PARP-inhibitot in breast cancer) ++ HOWEVER resistances & off-target effects have yet been poorly studied

### Toward an integrated map of genetic interactions in cancer cells (Rauscher et al.2018 )DOI: 10.15252/msb.20177656

* As basics for a new modelling method:
  + Previous description: statistical epitasis, defined as statistical deviation from the additive combination of two (gene) loci and how it affects the phenotype of interest
  + problematic, if data derives from various libraries or experimental protocols
  + new MINGLE method leads to interaction map sorted into genes with similar profiles and network molecules with similar functions
* Main Steps of the proposed method
  + available on <https://github.com/boutroslab/Supplemental-Material/tree/master/Rauscher_2017>
* Methodic elements of the model
  + considered alterations: gain or loss of copy events, somatic mutations and mRNA overexpression
  + gene-level correction and library batch effect determined with regression model with beta=0 and yi = y(crispr,i) - median(crispr,i) and a F-test (R-package sfsmisc) if the median crispr score is the same over all libraries (H0). Batch effect was defined as “phenotypes generated by one library differing significantly (FDR < 5%) from the observed median phenotype across all libraries”
  + control of normalization by Ward clustering/heatmap (hierachical clustering, do biologically related samples cluster closer than unrelated?)
  + comparison of control oncogenes towards gene knockout in the different libraries (2-sided student´s t-test); result was compared with literature
  + quantification of geneteic interactions with pi-score statistic
  + combinatorial testing of gene-gene-interactions: mixed model of regression analysis (fixed: alteration genotype, random: cell line)
  + gene similiratity network benchmarking and modeling (with pi-score and spearman; wasn´t able to understand the process completely)
* Key points of genetic interactions :
  + GANAB and PRKCSH as positive regulators of Wnt/beta-catenin signaling
  + RNF43 as negative regulator of the signaling pathway

References:

Berger, A.C., Korkut, A., Kanchi, R.S., Hegde, A.M., Lenoir, W., Liu, W., Liu, Y., Fan, H., Shen, H., Ravikumar, V., et al. (2018). A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. Cancer Cell 33, 690-705.e699. Ciriello, G., Gatza, Michael L., Beck, Andrew H., Wilkerson, Matthew D., Rhie, Suhn K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., et al. (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell 163, 506-519. The Cancer Genome Atlas, N., Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., et al. (2012). Comprehensive molecular portraits of human breast tumours. Nature 490, 61. Ashworth, A., Lord, C.J. & Reis-Filho, J.S., 2011. Genetic Interactions in Cancer Progression and Treatment. Cell, 145(1), S.30–38 Jerby-Arnon, Livnat, et al. “Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality.” Cell 158.5 (2014): 1199-120

### Possible Mutations for analysis:

1. PTEN
2. TP53
3. PIK3CA
4. MYC
5. GATA3
6. BRCA1/BRCA2
7. RB1
8. PARP
9. GANAB
10. PRKCSH
11. RNF43