

Lauren Saclauso's Project 2

Introduction

Scientific Question: Why do mutations in the BRCA1 gene lead to a 20% or higher risk of developing ovarian cancer compared to mutations in the BRCA2 gene?

BRCA1 & BRCA2:

The breast cancer 1 gene, also known as BRCA1, is a tumor suppressor gene found on chromosome 17 in humans. The breast cancer 2 gene, also known as BRCA2, is another tumor suppressor gene located on chromosome 13 (Mehrgou 2016). Both of these genes produce proteins that are involved in different steps of DNA damage response (DDR) and DNA repair (Roy 2011). There are 1863 amino acids in the BRCA1 protein and 3418 amino acids in the BRCA2 protein (Mehrgou 2016). Mutations in either of the genes can cause genetic instability and lead to a higher risk of developing breast and ovarian cancer, especially those that occur in the germline because they can continue to be passed down throughout generations (Welcsh 2001). It was found that BRCA1 mutation carriers have a 50% lifetime risk of developing ovarian cancer while BRCA2 mutation carriers have a 30% risk (Roy 2011) and that BRCA1 mutations become more common than BRCA2 mutations as the number of ovarian cancer cases within a family increase (Ramus 2009). In the 2000s, PALB2 was established as the physical and functional connection between BRCA1 and BRCA2 (Zhang 2009).

Scientific Hypothesis: If BRCA1 mutations are at least 20% more likely to lead to ovarian cancer than BRCA2 mutations, then there must be significant genetic and protein structural differences that account for increased risk.

I will be performing a pairwise sequence alignment to compare the protein sequences of BRCA1 and BRCA2 and then creating a dotplot to visualize the results of the alignment. The results from this will help me to identify any significant genetic differences between the two genes. Sequences will be downloaded from the NCBI database. Significant genetic differences could be used to explain why mutations in BRCA1 are more likely to cause ovarian cancer than mutations in BRCA2.

I will also be using structural bioinformatics to analyze the protein structures/complexes of BRCA1 and BRCA2. Additionally, I will attempt to determine how and where BRCA1 and BRCA2 interact by searching for binding site residues. The protein structures I analyze will be from the PDB and I will import them into PyMOL to visualize the protein structural differences and interactions. Identifying the protein structural differences could also reveal why BRCA1 mutations have higher penetrance than BRCA2 mutations. Observing how close their interactions may be can also account for their similar functions and explain why their mutations have comparable effects.

Loading In Packages

1. BiocManager: This package allows us to use packages from the Bioconductor project. Packages from this project have functions that can be used for statistical analysis of genomic data in R.
2. Biostrings: This is one of the packages from Bioconductor. It allows for manipulation of biological strings, such as DNA, RNA, and amino acids. I will be using this package to perform a global pairwise alignment of protein sequences.
3. seqinr: This package is used for retrieving and analyzing biological sequences. I will use seqinr to read in the fasta files of the protein sequences that I download from NCBI. I will also use the dotPlot function from this package to visualize the pairwise sequence alignment results.

4. Bio3D: The Bio3D package can be used for reading and writing biomolecular data types, analyzing sequences and structures, converting and manipulating data, plotting, and more. I will be using this package to read in coordinate files from PDB, look at internal structure, and identify interacting residues.
5. knitr: This package provides general-purpose tools for generating reports in R with Literate Programming techniques. It combines computing and reporting by incorporating code into texts documents. I will be using the `include_graphics` function to embed pictures of BRCA1 and BRCA2 complexes from PyMOL.
6. vembedr: This package is meant for embedding videos into R. Videos from YouTube, Vimeo, Box, and Microsoft stream can be embedded with their respective functions in this package. I will be embedding my PyMOL movie which I uploaded to YouTube.

Hide

```
# if (!require("BiocManager", quietly = TRUE))
#   install.packages("BiocManager")

# BiocManager::install("Biostrings")

library(BiocManager)
library(Biostrings)
library(seqinr)
```

Performing Bioinformatics Analysis

Pairwise Sequence Alignment: This bioinformatics method is used to compare two sequences by finding the optimal alignment between them. It then measures sequence similarity with an alignment score. High sequence similarity means that the sequences may share a functional, structural, or evolutionary relationship. We will be using the `pairwiseAlignment` function from the `Biostrings` package.

(<https://bioconductor.org/packages/devel/bioc/vignettes/Biostrings/inst/doc/PairwiseAlignments.pdf>)

(<https://bioconductor.org/packages/devel/bioc/vignettes/Biostrings/inst/doc/PairwiseAlignments.pdf>)

Hide

```
# Change global options so that entire AA sequence prints out - https://statisticsglobe.
com/reached-getoption-max-print-omitted-warning-in-r
options(max.print = 10000)

# Make sure bio3d package is detached if later part of code was already run since it als
o has a read.fasta function
# detach(package:bio3d, unload = TRUE)

# Read in fasta files and designate variables
# brca1_AA: https://www.ncbi.nlm.nih.gov/protein/AAC37594.1?report=fasta
brca1_AA <- read.fasta("BRCA1_AA.fasta", seqtype = "AA")
brca1_AA
```

\$AAC37594.1

[1] "M" "D" "L" "S" "A" "L" "R" "V" "E" "E" "V"
 [12] "Q" "N" "V" "I" "N" "A" "M" "Q" "K" "I" "L"
 [23] "E" "C" "P" "I" "C" "L" "E" "L" "I" "K" "E"
 [34] "P" "V" "S" "T" "K" "C" "D" "H" "I" "F" "C"
 [45] "K" "F" "C" "M" "L" "K" "L" "L" "N" "Q" "K"
 [56] "K" "G" "P" "S" "Q" "C" "P" "L" "C" "K" "N"
 [67] "D" "I" "T" "K" "R" "S" "L" "Q" "E" "S" "T"
 [78] "R" "F" "S" "Q" "L" "V" "E" "E" "L" "L" "K"
 [89] "I" "I" "C" "A" "F" "Q" "L" "D" "T" "G" "L"
 [100] "E" "Y" "A" "N" "S" "Y" "N" "F" "A" "K" "K"
 [111] "E" "N" "N" "S" "P" "E" "H" "L" "K" "D" "E"
 [122] "V" "S" "I" "I" "Q" "S" "M" "G" "Y" "R" "N"
 [133] "R" "A" "K" "R" "L" "L" "Q" "S" "E" "P" "E"
 [144] "N" "P" "S" "L" "Q" "E" "T" "S" "L" "S" "V"
 [155] "Q" "L" "S" "N" "L" "G" "T" "V" "R" "T" "L"
 [166] "R" "T" "K" "Q" "R" "I" "Q" "P" "Q" "K" "T"
 [177] "S" "V" "Y" "I" "E" "L" "G" "S" "D" "S" "S"
 [188] "E" "D" "T" "V" "N" "K" "A" "T" "Y" "C" "S"
 [199] "V" "G" "D" "Q" "E" "L" "L" "Q" "I" "T" "P"
 [210] "Q" "G" "T" "R" "D" "E" "I" "S" "L" "D" "S"
 [221] "A" "K" "K" "A" "A" "C" "E" "F" "S" "E" "T"
 [232] "D" "V" "T" "N" "T" "E" "H" "H" "Q" "P" "S"
 [243] "N" "N" "D" "L" "N" "T" "T" "E" "K" "R" "A"
 [254] "A" "E" "R" "H" "P" "E" "K" "Y" "Q" "G" "S"
 [265] "S" "V" "S" "N" "L" "H" "V" "E" "P" "C" "G"
 [276] "T" "N" "T" "H" "A" "S" "S" "L" "Q" "H" "E"
 [287] "N" "S" "S" "L" "L" "L" "T" "K" "D" "R" "M"
 [298] "N" "V" "E" "K" "A" "E" "F" "C" "N" "K" "S"
 [309] "K" "Q" "P" "G" "L" "A" "R" "S" "Q" "H" "N"
 [320] "R" "W" "A" "G" "S" "K" "E" "T" "C" "N" "D"
 [331] "R" "R" "T" "P" "S" "T" "E" "K" "K" "V" "D"
 [342] "L" "N" "A" "D" "P" "L" "C" "E" "R" "K" "E"
 [353] "W" "N" "K" "Q" "K" "L" "P" "C" "S" "E" "N"
 [364] "P" "R" "D" "T" "E" "D" "V" "P" "W" "I" "T"
 [375] "L" "N" "S" "S" "I" "Q" "K" "V" "N" "E" "W"
 [386] "F" "S" "R" "S" "D" "E" "L" "L" "G" "S" "D"
 [397] "D" "S" "H" "D" "G" "E" "S" "E" "S" "N" "A"
 [408] "K" "V" "A" "D" "V" "L" "D" "V" "L" "N" "E"
 [419] "V" "D" "E" "Y" "S" "G" "S" "S" "E" "K" "I"
 [430] "D" "L" "L" "A" "S" "D" "P" "H" "E" "A" "L"
 [441] "I" "C" "K" "S" "E" "R" "V" "H" "S" "K" "S"
 [452] "V" "E" "S" "N" "I" "E" "D" "K" "I" "F" "G"
 [463] "K" "T" "Y" "R" "K" "K" "A" "S" "L" "P" "N"
 [474] "L" "S" "H" "V" "T" "E" "N" "L" "I" "I" "G"
 [485] "A" "F" "V" "T" "E" "P" "Q" "I" "I" "Q" "E"
 [496] "R" "P" "L" "T" "N" "K" "L" "K" "R" "K" "R"
 [507] "R" "P" "T" "S" "G" "L" "H" "P" "E" "D" "F"
 [518] "I" "K" "K" "A" "D" "L" "A" "V" "Q" "K" "T"
 [529] "P" "E" "M" "I" "N" "Q" "G" "T" "N" "Q" "T"
 [540] "E" "Q" "N" "G" "Q" "V" "M" "N" "I" "T" "N"
 [551] "S" "G" "H" "E" "N" "K" "T" "K" "G" "D" "S"

[562] "I" "Q" "N" "E" "K" "N" "P" "N" "P" "I" "E"
 [573] "S" "L" "E" "K" "E" "S" "A" "F" "K" "T" "K"
 [584] "A" "E" "P" "I" "S" "S" "S" "I" "S" "N" "M"
 [595] "E" "L" "E" "L" "N" "I" "H" "N" "S" "K" "A"
 [606] "P" "K" "K" "N" "R" "L" "R" "R" "K" "S" "S"
 [617] "T" "R" "H" "I" "H" "A" "L" "E" "L" "V" "V"
 [628] "S" "R" "N" "L" "S" "P" "P" "N" "C" "T" "E"
 [639] "L" "Q" "I" "D" "S" "C" "S" "S" "S" "E" "E"
 [650] "I" "K" "K" "K" "K" "Y" "N" "Q" "M" "P" "V"
 [661] "R" "H" "S" "R" "N" "L" "Q" "L" "M" "E" "G"
 [672] "K" "E" "P" "A" "T" "G" "A" "K" "K" "S" "N"
 [683] "K" "P" "N" "E" "Q" "T" "S" "K" "R" "H" "D"
 [694] "S" "D" "T" "F" "P" "E" "L" "K" "L" "T" "N"
 [705] "A" "P" "G" "S" "F" "T" "K" "C" "S" "N" "T"
 [716] "S" "E" "L" "K" "E" "F" "V" "N" "P" "S" "L"
 [727] "P" "R" "E" "E" "K" "E" "E" "K" "L" "E" "T"
 [738] "V" "K" "V" "S" "N" "N" "A" "E" "D" "P" "K"
 [749] "D" "L" "M" "L" "S" "G" "E" "R" "V" "L" "Q"
 [760] "T" "E" "R" "S" "V" "E" "S" "S" "S" "I" "S"
 [771] "L" "V" "P" "G" "T" "D" "Y" "G" "T" "Q" "E"
 [782] "S" "I" "S" "L" "L" "E" "V" "S" "T" "L" "G"
 [793] "K" "A" "K" "T" "E" "P" "N" "K" "C" "V" "S"
 [804] "Q" "C" "A" "A" "F" "E" "N" "P" "K" "G" "L"
 [815] "I" "H" "G" "C" "S" "K" "D" "N" "R" "N" "D"
 [826] "T" "E" "G" "F" "K" "Y" "P" "L" "G" "H" "E"
 [837] "V" "N" "H" "S" "R" "E" "T" "S" "I" "E" "M"
 [848] "E" "E" "S" "E" "L" "D" "A" "Q" "Y" "L" "Q"
 [859] "N" "T" "F" "K" "V" "S" "K" "R" "Q" "S" "F"
 [870] "A" "P" "F" "S" "N" "P" "G" "N" "A" "E" "E"
 [881] "E" "C" "A" "T" "F" "S" "A" "H" "S" "G" "S"
 [892] "L" "K" "K" "Q" "S" "P" "K" "V" "T" "F" "E"
 [903] "C" "E" "Q" "K" "E" "E" "N" "Q" "G" "K" "N"
 [914] "E" "S" "N" "I" "K" "P" "V" "Q" "T" "V" "N"
 [925] "I" "T" "A" "G" "F" "P" "V" "V" "G" "Q" "K"
 [936] "D" "K" "P" "V" "D" "N" "A" "K" "C" "S" "I"
 [947] "K" "G" "G" "S" "R" "F" "C" "L" "S" "S" "Q"
 [958] "F" "R" "G" "N" "E" "T" "G" "L" "I" "T" "P"
 [969] "N" "K" "H" "G" "L" "L" "Q" "N" "P" "Y" "R"
 [980] "I" "P" "P" "L" "F" "P" "I" "K" "S" "F" "V"
 [991] "K" "T" "K" "C" "K" "K" "N" "L" "L" "E" "E"
 [1002] "N" "F" "E" "E" "H" "S" "M" "S" "P" "E" "R"
 [1013] "E" "M" "G" "N" "E" "N" "I" "P" "S" "T" "V"
 [1024] "S" "T" "I" "S" "R" "N" "N" "I" "R" "E" "N"
 [1035] "V" "F" "K" "E" "A" "S" "S" "S" "N" "I" "N"
 [1046] "E" "V" "G" "S" "S" "T" "N" "E" "V" "G" "S"
 [1057] "S" "I" "N" "E" "I" "G" "S" "S" "D" "E" "N"
 [1068] "I" "Q" "A" "E" "L" "G" "R" "N" "R" "G" "P"
 [1079] "K" "L" "N" "A" "M" "L" "R" "L" "G" "V" "L"
 [1090] "Q" "P" "E" "V" "Y" "K" "Q" "S" "L" "P" "G"
 [1101] "S" "N" "C" "K" "H" "P" "E" "I" "K" "K" "Q"
 [1112] "E" "Y" "E" "E" "V" "V" "Q" "T" "V" "N" "T"
 [1123] "D" "F" "S" "P" "Y" "L" "I" "S" "D" "N" "L"

[1134] "E" "Q" "P" "M" "G" "S" "S" "H" "A" "S" "Q"
 [1145] "V" "C" "S" "E" "T" "P" "D" "D" "L" "L" "D"
 [1156] "D" "G" "E" "I" "K" "E" "D" "T" "S" "F" "A"
 [1167] "E" "N" "D" "I" "K" "E" "S" "S" "A" "V" "F"
 [1178] "S" "K" "S" "V" "Q" "K" "G" "E" "L" "S" "R"
 [1189] "S" "P" "S" "P" "F" "T" "H" "T" "H" "L" "A"
 [1200] "Q" "G" "Y" "R" "R" "G" "A" "K" "K" "L" "E"
 [1211] "S" "S" "E" "E" "N" "L" "S" "S" "E" "D" "E"
 [1222] "E" "L" "P" "C" "F" "Q" "H" "L" "L" "F" "G"
 [1233] "K" "V" "N" "N" "I" "P" "S" "Q" "S" "T" "R"
 [1244] "H" "S" "T" "V" "A" "T" "E" "C" "L" "S" "K"
 [1255] "N" "T" "E" "E" "N" "L" "L" "S" "L" "K" "N"
 [1266] "S" "L" "N" "D" "C" "S" "N" "Q" "V" "I" "L"
 [1277] "A" "K" "A" "S" "Q" "E" "H" "H" "L" "S" "E"
 [1288] "E" "T" "K" "C" "S" "A" "S" "L" "F" "S" "S"
 [1299] "Q" "C" "S" "E" "L" "E" "D" "L" "T" "A" "N"
 [1310] "T" "N" "T" "Q" "D" "P" "F" "L" "I" "G" "S"
 [1321] "S" "K" "Q" "M" "R" "H" "Q" "S" "E" "S" "Q"
 [1332] "G" "V" "G" "L" "S" "D" "K" "E" "L" "V" "S"
 [1343] "D" "D" "E" "E" "R" "G" "T" "G" "L" "E" "E"
 [1354] "N" "N" "Q" "E" "E" "Q" "S" "M" "D" "S" "N"
 [1365] "L" "G" "E" "A" "A" "S" "G" "C" "E" "S" "E"
 [1376] "T" "S" "V" "S" "E" "D" "C" "S" "G" "L" "S"
 [1387] "S" "Q" "S" "D" "I" "L" "T" "T" "Q" "Q" "R"
 [1398] "D" "T" "M" "Q" "H" "N" "L" "I" "K" "L" "Q"
 [1409] "Q" "E" "M" "A" "E" "L" "E" "A" "V" "L" "E"
 [1420] "Q" "H" "G" "S" "Q" "P" "S" "N" "S" "Y" "P"
 [1431] "S" "I" "I" "S" "D" "S" "S" "A" "L" "E" "D"
 [1442] "L" "R" "N" "P" "E" "Q" "S" "T" "S" "E" "K"
 [1453] "A" "V" "L" "T" "S" "Q" "K" "S" "S" "E" "Y"
 [1464] "P" "I" "S" "Q" "N" "P" "E" "G" "L" "S" "A"
 [1475] "D" "K" "F" "E" "V" "S" "A" "D" "S" "S" "T"
 [1486] "S" "K" "N" "K" "E" "P" "G" "V" "E" "R" "S"
 [1497] "S" "P" "S" "K" "C" "P" "S" "L" "D" "D" "R"
 [1508] "W" "Y" "M" "H" "S" "C" "S" "G" "S" "L" "Q"
 [1519] "N" "R" "N" "Y" "P" "S" "Q" "E" "E" "L" "I"
 [1530] "K" "V" "V" "D" "V" "E" "E" "Q" "Q" "L" "E"
 [1541] "E" "S" "G" "P" "H" "D" "L" "T" "E" "T" "S"
 [1552] "Y" "L" "P" "R" "Q" "D" "L" "E" "G" "T" "P"
 [1563] "Y" "L" "E" "S" "G" "I" "S" "L" "F" "S" "D"
 [1574] "D" "P" "E" "S" "D" "P" "S" "E" "D" "R" "A"
 [1585] "P" "E" "S" "A" "R" "V" "G" "N" "I" "P" "S"
 [1596] "S" "T" "S" "A" "L" "K" "V" "P" "Q" "L" "K"
 [1607] "V" "A" "E" "S" "A" "Q" "S" "P" "A" "A" "A"
 [1618] "H" "T" "T" "D" "T" "A" "G" "Y" "N" "A" "M"
 [1629] "E" "E" "S" "V" "S" "R" "E" "K" "P" "E" "L"
 [1640] "T" "A" "S" "T" "E" "R" "V" "N" "K" "R" "M"
 [1651] "S" "M" "V" "V" "S" "G" "L" "T" "P" "E" "E"
 [1662] "F" "M" "L" "V" "Y" "K" "F" "A" "R" "K" "H"
 [1673] "H" "I" "T" "L" "T" "N" "L" "I" "T" "E" "E"
 [1684] "T" "T" "H" "V" "V" "M" "K" "T" "D" "A" "E"
 [1695] "F" "V" "C" "E" "R" "T" "L" "K" "Y" "F" "L"

```

[1706] "G" "I" "A" "G" "G" "K" "W" "V" "V" "S" "Y"
[1717] "F" "W" "V" "T" "Q" "S" "I" "K" "E" "R" "K"
[1728] "M" "L" "N" "E" "H" "D" "F" "E" "V" "R" "G"
[1739] "D" "V" "V" "N" "G" "R" "N" "H" "Q" "G" "P"
[1750] "K" "R" "A" "R" "E" "S" "Q" "D" "R" "K" "I"
[1761] "F" "R" "G" "L" "E" "I" "C" "C" "Y" "G" "P"
[1772] "F" "T" "N" "M" "P" "T" "D" "Q" "L" "E" "W"
[1783] "M" "V" "Q" "L" "C" "G" "A" "S" "V" "V" "K"
[1794] "E" "L" "S" "S" "F" "T" "L" "G" "T" "G" "V"
[1805] "H" "P" "I" "V" "V" "V" "Q" "P" "D" "A" "W"
[1816] "T" "E" "D" "N" "G" "F" "H" "A" "I" "G" "Q"
[1827] "M" "C" "E" "A" "P" "V" "V" "T" "R" "E" "W"
[1838] "V" "L" "D" "S" "V" "A" "L" "Y" "Q" "C" "Q"
[1849] "E" "L" "D" "T" "Y" "L" "I" "P" "Q" "I" "P"
[1860] "H" "S" "H" "Y"
attr(,"name")
[1] "AAC37594.1"
attr(,"Annot")
[1] ">AAC37594.1 BRCA1 [Homo sapiens]"
attr(,"class")
[1] "SeqFastaAA"

```

[Hide](#)

```

# brca2_AA: https://www.ncbi.nlm.nih.gov/protein/AAB07223.1?report=fasta
brca2_AA <- read.fasta("BRCA2_AA.fasta", seqtype = "AA")
brca2_AA

```

\$AAB07223.1

[1] "M" "P" "I" "G" "S" "K" "E" "R" "P" "T" "F"
 [12] "F" "E" "I" "F" "K" "T" "R" "C" "N" "K" "A"
 [23] "D" "L" "G" "P" "I" "S" "L" "N" "W" "F" "E"
 [34] "E" "L" "S" "S" "E" "A" "P" "P" "Y" "N" "S"
 [45] "E" "P" "A" "E" "E" "S" "E" "H" "K" "N" "N"
 [56] "N" "Y" "E" "P" "N" "L" "F" "K" "T" "P" "Q"
 [67] "R" "K" "P" "S" "Y" "N" "Q" "L" "A" "S" "T"
 [78] "P" "I" "I" "F" "K" "E" "Q" "G" "L" "T" "L"
 [89] "P" "L" "Y" "Q" "S" "P" "V" "K" "E" "L" "D"
 [100] "K" "F" "K" "L" "D" "L" "G" "R" "N" "V" "P"
 [111] "N" "S" "R" "H" "K" "S" "L" "R" "T" "V" "K"
 [122] "T" "K" "M" "D" "Q" "A" "D" "D" "V" "S" "C"
 [133] "P" "L" "L" "N" "S" "C" "L" "S" "E" "S" "P"
 [144] "V" "V" "L" "Q" "C" "T" "H" "V" "T" "P" "Q"
 [155] "R" "D" "K" "S" "V" "V" "C" "G" "S" "L" "F"
 [166] "H" "T" "P" "K" "F" "V" "K" "G" "R" "Q" "T"
 [177] "P" "K" "H" "I" "S" "E" "S" "L" "G" "A" "E"
 [188] "V" "D" "P" "D" "M" "S" "W" "S" "S" "S" "L"
 [199] "A" "T" "P" "P" "T" "L" "S" "S" "T" "V" "L"
 [210] "I" "V" "R" "N" "E" "E" "A" "S" "E" "T" "V"
 [221] "F" "P" "H" "D" "T" "T" "A" "N" "V" "K" "S"
 [232] "Y" "F" "S" "N" "H" "D" "E" "S" "L" "K" "K"
 [243] "N" "D" "R" "F" "I" "A" "S" "V" "T" "D" "S"
 [254] "E" "N" "T" "N" "Q" "R" "E" "A" "A" "S" "H"
 [265] "G" "F" "G" "K" "T" "S" "G" "N" "S" "F" "K"
 [276] "V" "N" "S" "C" "K" "D" "H" "I" "G" "K" "S"
 [287] "M" "P" "N" "V" "L" "E" "D" "E" "V" "Y" "E"
 [298] "T" "V" "V" "D" "T" "S" "E" "E" "D" "S" "F"
 [309] "S" "L" "C" "F" "S" "K" "C" "R" "T" "K" "N"
 [320] "L" "Q" "K" "V" "R" "T" "S" "K" "T" "R" "K"
 [331] "K" "I" "F" "H" "E" "A" "N" "A" "D" "E" "C"
 [342] "E" "K" "S" "K" "N" "Q" "V" "K" "E" "K" "Y"
 [353] "S" "F" "V" "S" "E" "V" "E" "P" "N" "D" "T"
 [364] "D" "P" "L" "D" "S" "N" "V" "A" "H" "Q" "K"
 [375] "P" "F" "E" "S" "G" "S" "D" "K" "I" "S" "K"
 [386] "E" "V" "V" "P" "S" "L" "A" "C" "E" "W" "S"
 [397] "Q" "L" "T" "L" "S" "G" "L" "N" "G" "A" "Q"
 [408] "M" "E" "K" "I" "P" "L" "L" "H" "I" "S" "S"
 [419] "C" "D" "Q" "N" "I" "S" "E" "K" "D" "L" "L"
 [430] "D" "T" "E" "N" "K" "R" "K" "K" "D" "F" "L"
 [441] "T" "S" "E" "N" "S" "L" "P" "R" "I" "S" "S"
 [452] "L" "P" "K" "S" "E" "K" "P" "L" "N" "E" "E"
 [463] "T" "V" "V" "N" "K" "R" "D" "E" "E" "Q" "H"
 [474] "L" "E" "S" "H" "T" "D" "C" "I" "L" "A" "V"
 [485] "K" "Q" "A" "I" "S" "G" "T" "S" "P" "V" "A"
 [496] "S" "S" "F" "Q" "G" "I" "K" "K" "S" "I" "F"
 [507] "R" "I" "R" "E" "S" "P" "K" "E" "T" "F" "N"
 [518] "A" "S" "F" "S" "G" "H" "M" "T" "D" "P" "N"
 [529] "F" "K" "K" "E" "T" "E" "A" "S" "E" "S" "G"
 [540] "L" "E" "I" "H" "T" "V" "C" "S" "Q" "K" "E"
 [551] "D" "S" "L" "C" "P" "N" "L" "I" "D" "N" "G"

[562] "S" "W" "P" "A" "T" "T" "T" "Q" "N" "S" "V"
 [573] "A" "L" "K" "N" "A" "G" "L" "I" "S" "T" "L"
 [584] "K" "K" "K" "T" "N" "K" "F" "I" "Y" "A" "I"
 [595] "H" "D" "E" "T" "F" "Y" "K" "G" "K" "K" "I"
 [606] "P" "K" "D" "Q" "K" "S" "E" "L" "I" "N" "C"
 [617] "S" "A" "Q" "F" "E" "A" "N" "A" "F" "E" "A"
 [628] "P" "L" "T" "F" "A" "N" "A" "D" "S" "G" "L"
 [639] "L" "H" "S" "S" "V" "K" "R" "S" "C" "S" "Q"
 [650] "N" "D" "S" "E" "E" "P" "T" "L" "S" "L" "T"
 [661] "S" "S" "F" "G" "T" "I" "L" "R" "K" "C" "S"
 [672] "R" "N" "E" "T" "C" "S" "N" "N" "T" "V" "I"
 [683] "S" "Q" "D" "L" "D" "Y" "K" "E" "A" "K" "C"
 [694] "N" "K" "E" "K" "L" "Q" "L" "F" "I" "T" "P"
 [705] "E" "A" "D" "S" "L" "S" "C" "L" "Q" "E" "G"
 [716] "Q" "C" "E" "N" "D" "P" "K" "S" "K" "K" "V"
 [727] "S" "D" "I" "K" "E" "E" "V" "L" "A" "A" "A"
 [738] "C" "H" "P" "V" "Q" "H" "S" "K" "V" "E" "Y"
 [749] "S" "D" "T" "D" "F" "Q" "S" "Q" "K" "S" "L"
 [760] "L" "Y" "D" "H" "E" "N" "A" "S" "T" "L" "I"
 [771] "L" "T" "P" "T" "S" "K" "D" "V" "L" "S" "N"
 [782] "L" "V" "M" "I" "S" "R" "G" "K" "E" "S" "Y"
 [793] "K" "M" "S" "D" "K" "L" "K" "G" "N" "N" "Y"
 [804] "E" "S" "D" "V" "E" "L" "T" "K" "N" "I" "P"
 [815] "M" "E" "K" "N" "Q" "D" "V" "C" "A" "L" "N"
 [826] "E" "N" "Y" "K" "N" "V" "E" "L" "L" "P" "P"
 [837] "E" "K" "Y" "M" "R" "V" "A" "S" "P" "S" "R"
 [848] "K" "V" "Q" "F" "N" "Q" "N" "T" "N" "L" "R"
 [859] "V" "I" "Q" "K" "N" "Q" "E" "E" "T" "T" "S"
 [870] "I" "S" "K" "I" "T" "V" "N" "P" "D" "S" "E"
 [881] "E" "L" "F" "S" "D" "N" "E" "N" "N" "F" "V"
 [892] "F" "Q" "V" "A" "N" "E" "R" "N" "N" "L" "A"
 [903] "L" "G" "N" "T" "K" "E" "L" "H" "E" "T" "D"
 [914] "L" "T" "C" "V" "N" "E" "P" "I" "F" "K" "N"
 [925] "S" "T" "M" "V" "L" "Y" "G" "D" "T" "G" "D"
 [936] "K" "Q" "A" "T" "Q" "V" "S" "I" "K" "K" "D"
 [947] "L" "V" "Y" "V" "L" "A" "E" "E" "N" "K" "N"
 [958] "S" "V" "K" "Q" "H" "I" "K" "M" "T" "L" "G"
 [969] "Q" "D" "L" "K" "S" "D" "I" "S" "L" "N" "I"
 [980] "D" "K" "I" "P" "E" "K" "N" "N" "D" "Y" "M"
 [991] "N" "K" "W" "A" "G" "L" "L" "G" "P" "I" "S"
 [1002] "N" "H" "S" "F" "G" "G" "S" "F" "R" "T" "A"
 [1013] "S" "N" "K" "E" "I" "K" "L" "S" "E" "H" "N"
 [1024] "I" "K" "K" "S" "K" "M" "F" "F" "K" "D" "I"
 [1035] "E" "E" "Q" "Y" "P" "T" "S" "L" "A" "C" "V"
 [1046] "E" "I" "V" "N" "T" "L" "A" "L" "D" "N" "Q"
 [1057] "K" "K" "L" "S" "K" "P" "Q" "S" "I" "N" "T"
 [1068] "V" "S" "A" "H" "L" "Q" "S" "S" "V" "V" "V"
 [1079] "S" "D" "C" "K" "N" "S" "H" "I" "T" "P" "Q"
 [1090] "M" "L" "F" "S" "K" "Q" "D" "F" "N" "S" "N"
 [1101] "H" "N" "L" "T" "P" "S" "Q" "K" "A" "E" "I"
 [1112] "T" "E" "L" "S" "T" "I" "L" "E" "E" "S" "G"
 [1123] "S" "Q" "F" "E" "F" "T" "Q" "F" "R" "K" "P"

[1134] "S" "Y" "I" "L" "Q" "K" "S" "T" "F" "E" "V"
[1145] "P" "E" "N" "Q" "M" "T" "I" "L" "K" "T" "T"
[1156] "S" "E" "E" "C" "R" "D" "A" "D" "L" "H" "V"
[1167] "I" "M" "N" "A" "P" "S" "I" "G" "Q" "V" "D"
[1178] "S" "S" "K" "Q" "F" "E" "G" "T" "V" "E" "I"
[1189] "K" "R" "K" "F" "A" "G" "L" "L" "K" "N" "D"
[1200] "C" "N" "K" "S" "A" "S" "G" "Y" "L" "T" "D"
[1211] "E" "N" "E" "V" "G" "F" "R" "G" "F" "Y" "S"
[1222] "A" "H" "G" "T" "K" "L" "N" "V" "S" "T" "E"
[1233] "A" "L" "Q" "K" "A" "V" "K" "L" "F" "S" "D"
[1244] "I" "E" "N" "I" "S" "E" "E" "T" "S" "A" "E"
[1255] "V" "H" "P" "I" "S" "L" "S" "S" "S" "K" "C"
[1266] "H" "D" "S" "V" "V" "S" "M" "F" "K" "I" "E"
[1277] "N" "H" "N" "D" "K" "T" "V" "S" "E" "K" "N"
[1288] "N" "K" "C" "Q" "L" "I" "L" "Q" "N" "N" "I"
[1299] "E" "M" "T" "T" "G" "T" "F" "V" "E" "E" "I"
[1310] "T" "E" "N" "Y" "K" "R" "N" "T" "E" "N" "E"
[1321] "D" "N" "K" "Y" "T" "A" "A" "S" "R" "N" "S"
[1332] "H" "N" "L" "E" "F" "D" "G" "S" "D" "S" "S"
[1343] "K" "N" "D" "T" "V" "C" "I" "H" "K" "D" "E"
[1354] "T" "D" "L" "L" "F" "T" "D" "Q" "H" "N" "I"
[1365] "C" "L" "K" "L" "S" "G" "Q" "F" "M" "K" "E"
[1376] "G" "N" "T" "Q" "I" "K" "E" "D" "L" "S" "D"
[1387] "L" "T" "F" "L" "E" "V" "A" "K" "A" "Q" "E"
[1398] "A" "C" "H" "G" "N" "T" "S" "N" "K" "E" "Q"
[1409] "L" "T" "A" "T" "K" "T" "E" "Q" "N" "I" "K"
[1420] "D" "F" "E" "T" "S" "D" "T" "F" "F" "Q" "T"
[1431] "A" "S" "G" "K" "N" "I" "S" "V" "A" "K" "E"
[1442] "S" "F" "N" "K" "I" "V" "N" "F" "F" "D" "Q"
[1453] "K" "P" "E" "E" "L" "H" "N" "F" "S" "L" "N"
[1464] "S" "E" "L" "H" "S" "D" "I" "R" "K" "N" "K"
[1475] "M" "D" "I" "L" "S" "Y" "E" "E" "T" "D" "I"
[1486] "V" "K" "H" "K" "I" "L" "K" "E" "S" "V" "P"
[1497] "V" "G" "T" "G" "N" "Q" "L" "V" "T" "F" "Q"
[1508] "G" "Q" "P" "E" "R" "D" "E" "K" "I" "K" "E"
[1519] "P" "T" "L" "L" "G" "F" "H" "T" "A" "S" "G"
[1530] "K" "K" "V" "K" "I" "A" "K" "E" "S" "L" "D"
[1541] "K" "V" "K" "N" "L" "F" "D" "E" "K" "E" "Q"
[1552] "G" "T" "S" "E" "I" "T" "S" "F" "S" "H" "Q"
[1563] "W" "A" "K" "T" "L" "K" "Y" "R" "E" "A" "C"
[1574] "K" "D" "L" "E" "L" "A" "C" "E" "T" "I" "E"
[1585] "I" "T" "A" "A" "P" "K" "C" "K" "E" "M" "Q"
[1596] "N" "S" "L" "N" "N" "D" "K" "N" "L" "V" "S"
[1607] "I" "E" "T" "V" "V" "P" "P" "K" "L" "L" "S"
[1618] "D" "N" "L" "C" "R" "Q" "T" "E" "N" "L" "K"
[1629] "T" "S" "K" "S" "I" "F" "L" "K" "V" "K" "V"
[1640] "H" "E" "N" "V" "E" "K" "E" "T" "A" "K" "S"
[1651] "P" "A" "T" "C" "Y" "T" "N" "Q" "S" "P" "Y"
[1662] "S" "V" "I" "E" "N" "S" "A" "L" "A" "F" "Y"
[1673] "T" "S" "C" "S" "R" "K" "T" "S" "V" "S" "Q"
[1684] "T" "S" "L" "L" "E" "A" "K" "K" "W" "L" "R"
[1695] "E" "G" "I" "F" "D" "G" "Q" "P" "E" "R" "I"

[1706] "N" "T" "A" "D" "Y" "V" "G" "N" "Y" "L" "Y"
 [1717] "E" "N" "N" "S" "N" "S" "T" "I" "A" "E" "N"
 [1728] "D" "K" "N" "H" "L" "S" "E" "K" "Q" "D" "T"
 [1739] "Y" "L" "S" "N" "S" "S" "M" "S" "N" "S" "Y"
 [1750] "S" "Y" "H" "S" "D" "E" "V" "Y" "N" "D" "S"
 [1761] "G" "Y" "L" "S" "K" "N" "K" "L" "D" "S" "G"
 [1772] "I" "E" "P" "V" "L" "K" "N" "V" "E" "D" "Q"
 [1783] "K" "N" "T" "S" "F" "S" "K" "V" "I" "S" "N"
 [1794] "V" "K" "D" "A" "N" "A" "Y" "P" "Q" "T" "V"
 [1805] "N" "E" "D" "I" "C" "V" "E" "E" "L" "V" "T"
 [1816] "S" "S" "S" "P" "C" "K" "N" "K" "N" "A" "A"
 [1827] "I" "K" "L" "S" "I" "S" "N" "S" "N" "N" "F"
 [1838] "E" "V" "G" "P" "P" "A" "F" "R" "I" "A" "S"
 [1849] "G" "K" "I" "V" "C" "V" "S" "H" "E" "T" "I"
 [1860] "K" "K" "V" "K" "D" "I" "F" "T" "D" "S" "F"
 [1871] "S" "K" "V" "I" "K" "E" "N" "N" "E" "N" "K"
 [1882] "S" "K" "I" "C" "Q" "T" "K" "I" "M" "A" "G"
 [1893] "C" "Y" "E" "A" "L" "D" "D" "S" "E" "D" "I"
 [1904] "L" "H" "N" "S" "L" "D" "N" "D" "E" "C" "S"
 [1915] "T" "H" "S" "H" "K" "V" "F" "A" "D" "I" "Q"
 [1926] "S" "E" "E" "I" "L" "Q" "H" "N" "Q" "N" "M"
 [1937] "S" "G" "L" "E" "K" "V" "S" "K" "I" "S" "P"
 [1948] "C" "D" "V" "S" "L" "E" "T" "S" "D" "I" "C"
 [1959] "K" "C" "S" "I" "G" "K" "L" "H" "K" "S" "V"
 [1970] "S" "S" "A" "N" "T" "C" "G" "I" "F" "S" "T"
 [1981] "A" "S" "G" "K" "S" "V" "Q" "V" "S" "D" "A"
 [1992] "S" "L" "Q" "N" "A" "R" "Q" "V" "F" "S" "E"
 [2003] "I" "E" "D" "S" "T" "K" "Q" "V" "F" "S" "K"
 [2014] "V" "L" "F" "K" "S" "N" "E" "H" "S" "D" "Q"
 [2025] "L" "T" "R" "E" "E" "N" "T" "A" "I" "R" "T"
 [2036] "P" "E" "H" "L" "I" "S" "Q" "K" "G" "F" "S"
 [2047] "Y" "N" "V" "V" "N" "S" "S" "A" "F" "S" "G"
 [2058] "F" "S" "T" "A" "S" "G" "K" "Q" "V" "S" "I"
 [2069] "L" "E" "S" "S" "L" "H" "K" "V" "K" "G" "V"
 [2080] "L" "E" "E" "F" "D" "L" "I" "R" "T" "E" "H"
 [2091] "S" "L" "H" "Y" "S" "P" "T" "S" "R" "Q" "N"
 [2102] "V" "S" "K" "I" "L" "P" "R" "V" "D" "K" "R"
 [2113] "N" "P" "E" "H" "C" "V" "N" "S" "E" "M" "E"
 [2124] "K" "T" "C" "S" "K" "E" "F" "K" "L" "S" "N"
 [2135] "N" "L" "N" "V" "E" "G" "G" "S" "S" "E" "N"
 [2146] "N" "H" "S" "I" "K" "V" "S" "P" "Y" "L" "S"
 [2157] "Q" "F" "Q" "Q" "D" "K" "Q" "Q" "L" "V" "L"
 [2168] "G" "T" "K" "V" "S" "L" "V" "E" "N" "I" "H"
 [2179] "V" "L" "G" "K" "E" "Q" "A" "S" "P" "K" "N"
 [2190] "V" "K" "M" "E" "I" "G" "K" "T" "E" "T" "F"
 [2201] "S" "D" "V" "P" "V" "K" "T" "N" "I" "E" "V"
 [2212] "C" "S" "T" "Y" "S" "K" "D" "S" "E" "N" "Y"
 [2223] "F" "E" "T" "E" "A" "V" "E" "I" "A" "K" "A"
 [2234] "F" "M" "E" "D" "D" "E" "L" "T" "D" "S" "K"
 [2245] "L" "P" "S" "H" "A" "T" "H" "S" "L" "F" "T"
 [2256] "C" "P" "E" "N" "E" "E" "M" "V" "L" "S" "N"
 [2267] "S" "R" "I" "G" "K" "R" "R" "G" "E" "P" "L"

[2278] "I" "L" "V" "G" "E" "P" "S" "I" "K" "R" "N"
 [2289] "L" "L" "N" "E" "F" "D" "R" "I" "I" "E" "N"
 [2300] "Q" "E" "K" "S" "L" "K" "A" "S" "K" "S" "T"
 [2311] "P" "D" "G" "T" "I" "K" "D" "R" "R" "L" "F"
 [2322] "M" "H" "H" "V" "S" "L" "E" "P" "I" "T" "C"
 [2333] "V" "P" "F" "R" "T" "T" "K" "E" "R" "Q" "E"
 [2344] "I" "Q" "N" "P" "N" "F" "T" "A" "P" "G" "Q"
 [2355] "E" "F" "L" "S" "K" "S" "H" "L" "Y" "E" "H"
 [2366] "L" "T" "L" "E" "K" "S" "S" "S" "N" "L" "A"
 [2377] "V" "S" "G" "H" "P" "F" "Y" "Q" "V" "S" "A"
 [2388] "T" "R" "N" "E" "K" "M" "R" "H" "L" "I" "T"
 [2399] "T" "G" "R" "P" "T" "K" "V" "F" "V" "P" "P"
 [2410] "F" "K" "T" "K" "S" "H" "F" "H" "R" "V" "E"
 [2421] "Q" "C" "V" "R" "N" "I" "N" "L" "E" "E" "N"
 [2432] "R" "Q" "K" "Q" "N" "I" "D" "G" "H" "G" "S"
 [2443] "D" "D" "S" "K" "N" "K" "I" "N" "D" "N" "E"
 [2454] "I" "H" "Q" "F" "N" "K" "N" "N" "S" "N" "Q"
 [2465] "A" "A" "A" "V" "T" "F" "T" "K" "C" "E" "E"
 [2476] "E" "P" "L" "D" "L" "I" "T" "S" "L" "Q" "N"
 [2487] "A" "R" "D" "I" "Q" "D" "M" "R" "I" "K" "K"
 [2498] "K" "Q" "R" "Q" "R" "V" "F" "P" "Q" "P" "G"
 [2509] "S" "L" "Y" "L" "A" "K" "T" "S" "T" "L" "P"
 [2520] "R" "I" "S" "L" "K" "A" "A" "V" "G" "G" "Q"
 [2531] "V" "P" "S" "A" "C" "S" "H" "K" "Q" "L" "Y"
 [2542] "T" "Y" "G" "V" "S" "K" "H" "C" "I" "K" "I"
 [2553] "N" "S" "K" "N" "A" "E" "S" "F" "Q" "F" "H"
 [2564] "T" "E" "D" "Y" "F" "G" "K" "E" "S" "L" "W"
 [2575] "T" "G" "K" "G" "I" "Q" "L" "A" "D" "G" "G"
 [2586] "W" "L" "I" "P" "S" "N" "D" "G" "K" "A" "G"
 [2597] "K" "E" "E" "F" "Y" "R" "A" "L" "C" "D" "T"
 [2608] "P" "G" "V" "D" "P" "K" "L" "I" "S" "R" "I"
 [2619] "W" "V" "Y" "N" "H" "Y" "R" "W" "I" "I" "W"
 [2630] "K" "L" "A" "A" "M" "E" "C" "A" "F" "P" "K"
 [2641] "E" "F" "A" "N" "R" "C" "L" "S" "P" "E" "R"
 [2652] "V" "L" "L" "Q" "L" "K" "Y" "R" "Y" "D" "T"
 [2663] "E" "I" "D" "R" "S" "R" "R" "S" "A" "I" "K"
 [2674] "K" "I" "M" "E" "R" "D" "D" "T" "A" "A" "K"
 [2685] "T" "L" "V" "L" "C" "V" "S" "D" "I" "I" "S"
 [2696] "L" "S" "A" "N" "I" "S" "E" "T" "S" "S" "N"
 [2707] "K" "T" "S" "S" "A" "D" "T" "Q" "K" "V" "A"
 [2718] "I" "I" "E" "L" "T" "D" "G" "W" "Y" "A" "V"
 [2729] "K" "A" "Q" "L" "D" "P" "P" "L" "L" "A" "V"
 [2740] "L" "K" "N" "G" "R" "L" "T" "V" "G" "Q" "K"
 [2751] "I" "I" "L" "H" "G" "A" "E" "L" "V" "G" "S"
 [2762] "P" "D" "A" "C" "T" "P" "L" "E" "A" "P" "E"
 [2773] "S" "L" "M" "L" "K" "I" "S" "A" "N" "S" "T"
 [2784] "R" "P" "A" "R" "W" "Y" "T" "K" "L" "G" "F"
 [2795] "F" "P" "D" "P" "R" "P" "F" "P" "L" "P" "L"
 [2806] "S" "S" "L" "F" "S" "D" "G" "G" "N" "V" "G"
 [2817] "C" "V" "D" "V" "I" "I" "Q" "R" "A" "Y" "P"
 [2828] "I" "Q" "W" "M" "E" "K" "T" "S" "S" "G" "L"
 [2839] "Y" "I" "F" "R" "N" "E" "R" "E" "E" "E" "K"

[2850] "E" "A" "A" "K" "Y" "V" "E" "A" "Q" "Q" "K"
 [2861] "R" "L" "E" "A" "L" "F" "T" "K" "I" "Q" "E"
 [2872] "E" "F" "E" "E" "H" "E" "E" "N" "T" "T" "K"
 [2883] "P" "Y" "L" "P" "S" "R" "A" "L" "T" "R" "Q"
 [2894] "Q" "V" "R" "A" "L" "Q" "D" "G" "A" "E" "L"
 [2905] "Y" "E" "A" "V" "K" "N" "A" "A" "D" "P" "A"
 [2916] "Y" "L" "E" "G" "Y" "F" "S" "E" "E" "Q" "L"
 [2927] "R" "A" "L" "N" "N" "H" "R" "Q" "M" "L" "N"
 [2938] "D" "K" "K" "Q" "A" "Q" "I" "Q" "L" "E" "I"
 [2949] "R" "K" "A" "M" "E" "S" "A" "E" "Q" "K" "E"
 [2960] "Q" "G" "L" "S" "R" "D" "V" "T" "T" "V" "W"
 [2971] "K" "L" "R" "I" "V" "S" "Y" "S" "K" "K" "E"
 [2982] "K" "D" "S" "V" "I" "L" "S" "I" "W" "R" "P"
 [2993] "S" "S" "D" "L" "Y" "S" "L" "L" "T" "E" "G"
 [3004] "K" "R" "Y" "R" "I" "Y" "H" "L" "A" "T" "S"
 [3015] "K" "S" "K" "S" "K" "S" "E" "R" "A" "N" "I"
 [3026] "Q" "L" "A" "A" "T" "K" "K" "T" "Q" "Y" "Q"
 [3037] "Q" "L" "P" "V" "S" "D" "E" "I" "L" "F" "Q"
 [3048] "I" "Y" "Q" "P" "R" "E" "P" "L" "H" "F" "S"
 [3059] "K" "F" "L" "D" "P" "D" "F" "Q" "P" "S" "C"
 [3070] "S" "E" "V" "D" "L" "I" "G" "F" "V" "V" "S"
 [3081] "V" "V" "K" "K" "T" "G" "L" "A" "P" "F" "V"
 [3092] "Y" "L" "S" "D" "E" "C" "Y" "N" "L" "L" "A"
 [3103] "I" "K" "F" "W" "I" "D" "L" "N" "E" "D" "I"
 [3114] "I" "K" "P" "H" "M" "L" "I" "A" "A" "S" "N"
 [3125] "L" "Q" "W" "R" "P" "E" "S" "K" "S" "G" "L"
 [3136] "L" "T" "L" "F" "A" "G" "D" "F" "S" "V" "F"
 [3147] "S" "A" "S" "P" "K" "E" "G" "H" "F" "Q" "E"
 [3158] "T" "F" "N" "K" "M" "K" "N" "T" "V" "E" "N"
 [3169] "I" "D" "I" "L" "C" "N" "E" "A" "E" "N" "K"
 [3180] "L" "M" "H" "I" "L" "H" "A" "N" "D" "P" "K"
 [3191] "W" "S" "T" "P" "T" "K" "D" "C" "T" "S" "G"
 [3202] "P" "Y" "T" "A" "Q" "I" "I" "P" "G" "T" "G"
 [3213] "N" "K" "L" "L" "M" "S" "S" "P" "N" "C" "E"
 [3224] "I" "Y" "Y" "Q" "S" "P" "L" "S" "L" "C" "M"
 [3235] "A" "K" "R" "K" "S" "V" "S" "T" "P" "V" "S"
 [3246] "A" "Q" "M" "T" "S" "K" "S" "C" "K" "G" "E"
 [3257] "K" "E" "I" "D" "D" "Q" "K" "N" "C" "K" "K"
 [3268] "R" "R" "A" "L" "D" "F" "L" "S" "R" "L" "P"
 [3279] "L" "P" "P" "P" "V" "S" "P" "I" "C" "T" "F"
 [3290] "V" "S" "P" "A" "A" "Q" "K" "A" "F" "Q" "P"
 [3301] "P" "R" "S" "C" "G" "T" "K" "Y" "E" "T" "P"
 [3312] "I" "K" "K" "K" "E" "L" "N" "S" "P" "Q" "M"
 [3323] "T" "P" "F" "K" "K" "F" "N" "E" "I" "S" "L"
 [3334] "L" "E" "S" "N" "S" "I" "A" "D" "E" "E" "L"
 [3345] "A" "L" "I" "N" "T" "Q" "A" "L" "L" "S" "G"
 [3356] "S" "T" "G" "E" "K" "Q" "F" "I" "S" "V" "S"
 [3367] "E" "S" "T" "R" "T" "A" "P" "T" "S" "S" "E"
 [3378] "D" "Y" "L" "R" "L" "K" "R" "R" "C" "T" "T"
 [3389] "S" "L" "I" "K" "E" "Q" "E" "S" "S" "Q" "A"
 [3400] "S" "T" "E" "E" "C" "E" "K" "N" "K" "Q" "D"
 [3411] "T" "I" "T" "T" "K" "K" "Y" "I"

```
attr(,"name")
[1] "AAB07223.1"
attr(,"Annot")
[1] ">AAB07223.1 BRCA2 [Homo sapiens]"
attr(,"class")
[1] "SeqFastaAA"
```

Hide

```
#class(brca1_AA)
#class(brca2_AA)
# Convert from list to character strings to prepare for pairwise sequence alignment
brca1 <- toString(brca1_AA)
brca2 <- toString(brca2_AA)
#class(brca1)
#class(brca2)

# Find optimal global alignment between protein sequences
brca_alignment <- pairwiseAlignment(pattern = brca1, subject = brca2, type = "global")

# Print optimal global alignment and score
brca_alignment
```

```
Global PairwiseAlignmentsSingleSubject (1 of 1)
pattern: c("M", "D", "L", "-...S", "H", "Y-----")
subject: c("M", "P", "I", "G..."K", "K", "Y", "I")
score: -2500.069
```

Structural Bioinformatics: This method is used to determine, analyze, and predict 3D structures of biomolecules. These biomolecules can be DNA, RNA, or proteins. The information obtained from this method can help us to strengthen our understanding of the relationship between structure and function. We will be using the `read.pdb` function from the `Bio3D` package to read in the PDB coordinate files for the crystal structure of the BARD1 BRCT domains (2NTE) and the crystal structure of a PALB2 / BRCA2 complex (3EUZ). Additionally, we will determine if and where the two structures have interacting residues with the `binding.site` function which is also from `Bio3D`. (http://thegrantlab.org/bio3d_v2/html/read.pdb.html (http://thegrantlab.org/bio3d_v2/html/read.pdb.html), http://thegrantlab.org/bio3d_v2/html/binding.site.html (http://thegrantlab.org/bio3d_v2/html/binding.site.html))

Hide

```
# Call bio3d package after pairwise alignment to avoid mixing up read.fasta functions
library(bio3d)
```

```
Attaching package: 'bio3d'
```

```
The following objects are masked from 'package:seqinr':
```

```
consensus, read.fasta, write.fasta
```

```
The following object is masked from 'package:Biostrings':
```

```
mask
```

```
The following object is masked from 'package:IRanges':
```

```
trim
```

[Hide](#)

```
# Read PDB files and print composition summaries  
BARD1_BRCT <- read.pdb("2nte")
```

```
Note: Accessing on-line PDB file
```

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE) :  
  /var/folders/8x/nkwm9lss4c9frr6lz75vftcr0000gn/T//RtmpffjZWe/2nte.pdb exists. Skipping  
  download
```

```
PDB has ALT records, taking A only, rm.alt=TRUE
```

[Hide](#)

```
BARD1_BRCT
```

```
Call: read.pdb(file = "2nte")
```

```
Total Models#: 1
```

```
Total Atoms#: 3652, XYZs#: 10956 Chains#: 2 (values: A B)
```

```
Protein Atoms#: 3382 (residues/Calpha atoms#: 419)
```

```
Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)
```

```
Non-protein/nucleic Atoms#: 270 (residues: 247)
```

```
Non-protein/nucleic resid values: [ CL (3), EDO (5), HOH (237), SO4 (2) ]
```

```
Protein sequence:
```

```
GPLVLIGSGLSSEQQKMLSELAVILKAKKYTEFDSTVTHVVVPGDAVQSTLKCMLGILNG
CWILKFEWVKACLRKRVCEQEEKYEIPEGPRRSRLNREQLLPKLFDCYFYLGWTFKHHP
KDNLIKLVTAGGGQILSRKPKPDSVDVTQTINTVAYHARPDSQRFCTQYIIYEDLCNYHP
ERVQRQGVWKAPSSWFIDCVMSFELLPLDSPLVLIGSGLSSEQQK...<cut>...PLDS
```

```
+ attr: atom, xyz, seqres, helix, sheet,
      calpha, remark, call
```

[Hide](#)

```
PALB2_BRCA2 <- read.pdb("3euz")
```

```
Note: Accessing on-line PDB file
```

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE) :
  /var/folders/8x/nkwm9lss4c9frr6lz75vftcr0000gn/T//RtmpffjzWe/3euz.pdb exists. Skipping
  download
```

```
PDB has ALT records, taking A only, rm.alt=TRUE
```

[Hide](#)

```
PALB2_BRCA2
```

```
Call: read.pdb(file = "3euz")
```

```
Total Models#: 1
```

```
Total Atoms#: 2095, XYZs#: 6285 Chains#: 2 (values: A B)
```

```
Protein Atoms#: 1848 (residues/Calpha atoms#: 239)
```

```
Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)
```

```
Non-protein/nucleic Atoms#: 247 (residues: 243)
```

```
Non-protein/nucleic resid values: [ DMF (1), HOH (242) ]
```

```
Protein sequence:
```

```
ETAAAKFERQHMDSSSSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAVCSQKNVAC  
KNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIIVACEGNPYVPVHFDASVK  
ETAAAKFERQHMDSSTSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAVCSQKNVAC  
KNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIIVAC...<cut>...DASV
```

```
+ attr: atom, xyz, seqres, helix, sheet,  
      calpha, remark, call
```

[Hide](#)

```
# Display internal structures  
str(BARD1_BRCT)
```


List of 8

```
$ atom : 'data.frame': 3652 obs. of 16 variables:
..$ type : chr [1:3652] "ATOM" "ATOM" "ATOM" "ATOM" ...
..$ eleno : int [1:3652] 1 2 3 4 5 6 7 8 9 10 ...
..$ elety : chr [1:3652] "N" "CA" "C" "O" ...
..$ alt : chr [1:3652] NA NA NA NA ...
..$ resid : chr [1:3652] "GLY" "GLY" "GLY" "GLY" ...
..$ chain : chr [1:3652] "A" "A" "A" "A" ...
..$ resno : int [1:3652] 568 568 568 568 569 569 569 569 569 ...
..$ insert: chr [1:3652] NA NA NA NA ...
..$ x : num [1:3652] 26.4 27.5 27 25.9 27.9 ...
..$ y : num [1:3652] 0.267 1.283 2.736 2.954 3.718 ...
..$ z : num [1:3652] 45.7 45.5 45.7 46.1 45.5 ...
..$ o : num [1:3652] 1 1 1 1 1 1 1 1 1 1 ...
..$ b : num [1:3652] 67.6 67 66.9 67.3 65.7 ...
..$ segid : chr [1:3652] NA NA NA NA ...
..$ elesy : chr [1:3652] "N" "C" "C" "O" ...
..$ charge: chr [1:3652] NA NA NA NA ...
$ xyz : 'xyz' num [1, 1:10956] 26.364 0.267 45.718 27.462 1.283 ...
$ seqres: Named chr [1:420] "GLY" "PRO" "LEU" "VAL" ...
..- attr(*, "names")= chr [1:420] "A" "A" "A" "A" ...
$ helix :List of 4
..$ start: Named num [1:18] 578 617 633 647 655 687 708 728 760 578 ...
.. ..- attr(*, "names")= chr [1:18] "" "" "" "" ...
..$ end : Named num [1:18] 592 626 643 651 666 698 716 732 770 592 ...
.. ..- attr(*, "names")= chr [1:18] "" "" "" "" ...
..$ chain: chr [1:18] "A" "A" "A" "A" ...
..$ type : chr [1:18] "1" "1" "1" "5" ...
$ sheet :List of 4
..$ start: Named num [1:18] 595 571 606 629 701 676 735 755 751 595 ...
.. ..- attr(*, "names")= chr [1:18] "" "" "" "" ...
..$ end : Named num [1:18] 597 574 609 632 703 679 739 759 752 596 ...
.. ..- attr(*, "names")= chr [1:18] "" "" "" "" ...
..$ chain: chr [1:18] "A" "A" "A" "A" ...
..$ sense: chr [1:18] "0" "1" "1" "1" ...
$ calpha: logi [1:3652] FALSE TRUE FALSE FALSE FALSE TRUE ...
$ remark:List of 1
..$ biomat:List of 4
.. ..$ num : int 2
.. ..$ chain :List of 2
.. .. ..$ : chr "A"
.. .. ..$ : chr "B"
.. ..$ mat :List of 2
.. .. ..$ :List of 1
.. .. .. ..$ A: num [1:3, 1:4] 1 0 0 0 1 0 0 0 1 0 ...
.. .. .. ..$ :List of 1
.. .. .. ..$ B: num [1:3, 1:4] 1 0 0 0 1 0 0 0 1 0 ...
.. ..$ method: chr [1:2] "AUTHOR" "AUTHOR"
$ call : language read.pdb(file = "2nte")
- attr(*, "class")= chr [1:2] "pdb" "sse"
```

[Hide](#)

```
str(PALB2_BRCA2)
```

List of 8

```
$ atom : 'data.frame': 2095 obs. of 16 variables:
..$ type : chr [1:2095] "ATOM" "ATOM" "ATOM" "ATOM" ...
..$ eleno : int [1:2095] 1 2 3 4 5 6 7 8 9 10 ...
..$ elety : chr [1:2095] "N" "CA" "C" "O" ...
..$ alt : chr [1:2095] NA NA NA NA ...
..$ resid : chr [1:2095] "GLU" "GLU" "GLU" "GLU" ...
..$ chain : chr [1:2095] "A" "A" "A" "A" ...
..$ resno : int [1:2095] 2 2 2 2 2 2 2 2 2 3 ...
..$ insert: chr [1:2095] NA NA NA NA ...
..$ x : num [1:2095] -86.3 -85.4 -85.4 -86.5 -85.9 ...
..$ y : num [1:2095] 16.6 15.4 14.9 14.6 14.3 ...
..$ z : num [1:2095] 26.1 26 24.6 24 26.9 ...
..$ o : num [1:2095] 1 1 1 1 1 1 1 1 1 1 ...
..$ b : num [1:2095] 44.8 43.7 42.1 42.3 44.8 ...
..$ segid : chr [1:2095] NA NA NA NA ...
..$ elesy : chr [1:2095] "N" "C" "C" "O" ...
..$ charge: chr [1:2095] NA NA NA NA ...
$ xyz : 'xyz' num [1, 1:6285] -86.3 16.6 26.1 -85.4 15.4 ...
$ seqres: Named chr [1:248] "LYS" "GLU" "THR" "ALA" ...
..- attr(*, "names")= chr [1:248] "A" "A" "A" "A" ...
$ helix :List of 4
..$ start: Named num [1:8] 3 24 50 57 3 24 50 57
.. ..- attr(*, "names")= chr [1:8] "" "" "" "" ...
..$ end : Named num [1:8] 13 33 56 60 13 33 56 60
.. ..- attr(*, "names")= chr [1:8] "" "" "" "" ...
..$ chain: chr [1:8] "A" "A" "A" "A" ...
..$ type : chr [1:8] "1" "1" "1" "5" ...
$ sheet :List of 4
..$ start: Named num [1:12] 43 79 97 72 61 116 43 79 97 72 ...
.. ..- attr(*, "names")= chr [1:12] "" "" "" "" ...
..$ end : Named num [1:12] 47 86 111 74 63 124 47 86 111 74 ...
.. ..- attr(*, "names")= chr [1:12] "" "" "" "" ...
..$ chain: chr [1:12] "A" "A" "A" "A" ...
..$ sense: chr [1:12] "0" "-1" "-1" "-1" ...
$ calpha: logi [1:2095] FALSE TRUE FALSE FALSE FALSE FALSE ...
$ remark:List of 1
..$ biomat:List of 4
.. ..$ num : int 2
.. ..$ chain :List of 2
.. .. ..$ : chr "A"
.. .. ..$ : chr "B"
.. ..$ mat :List of 2
.. .. ..$ :List of 1
.. .. .. ..$ A: num [1:3, 1:4] 1 0 0 0 1 0 0 0 1 0 ...
.. .. .. ..$ :List of 1
.. .. .. ..$ B: num [1:3, 1:4] 1 0 0 0 1 0 0 0 1 0 ...
.. ..$ method: chr [1:2] "AUTHOR" "AUTHOR"
$ call : language read.pdb(file = "3euz")
- attr(*, "class")= chr [1:2] "pdb" "sse"
```

Hide

```
# Determine and print interaction(s) between BARD1_BRCT & PALB2_BRCA with a cutoff of 50
# No interacting residues appeared until a cutoff of 42
BSR <- binding.site(BARD1_BRCT, PALB2_BRCA2, cutoff = 50, hydrogens = TRUE, byres = TRUE
, verbose = TRUE)
```

```
.. 00003652 atom(s) from 'string' selection
.. 00003652 atom(s) in final combined selection

.. 00002095 atom(s) from 'string' selection
.. 00002095 atom(s) in final combined selection
```

Hide

BSR

\$inds

Call: NULL

```
Atom Indices#: 87 ($atom)
XYZ Indices#: 261 ($xyz)
```

+ attr: atom, xyz

\$resnames

```
[1] "GLU 655 (A)" "ARG 725 (A)" "PRO 726 (A)"
[4] "ASP 727 (A)" "SER 728 (A)" "GLU 771 (A)"
[7] "LEU 773 (A)" "PRO 774 (A)" "ASP 776 (A)"
[10] "SER 777 (A)" "HOH 88 (A)" "HOH 203 (A)"
[13] "HOH 204 (A)" "HOH 209 (A)" "HOH 211 (A)"
[16] "HOH 227 (A)" "HOH 276 (A)"
```

\$resno

```
[1] 655 725 726 727 728 771 773 774 776 777 88
[12] 203 204 209 211 227 276
```

\$chain

```
[1] "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A"
[12] "A" "A" "A" "A" "A" "A"
```

\$call

```
binding.site(a = BARD1_BRCT, b = PALB2_BRCA2, cutoff = 50, hydrogens = TRUE,
byres = TRUE, verbose = TRUE)
```

Plotting The Results

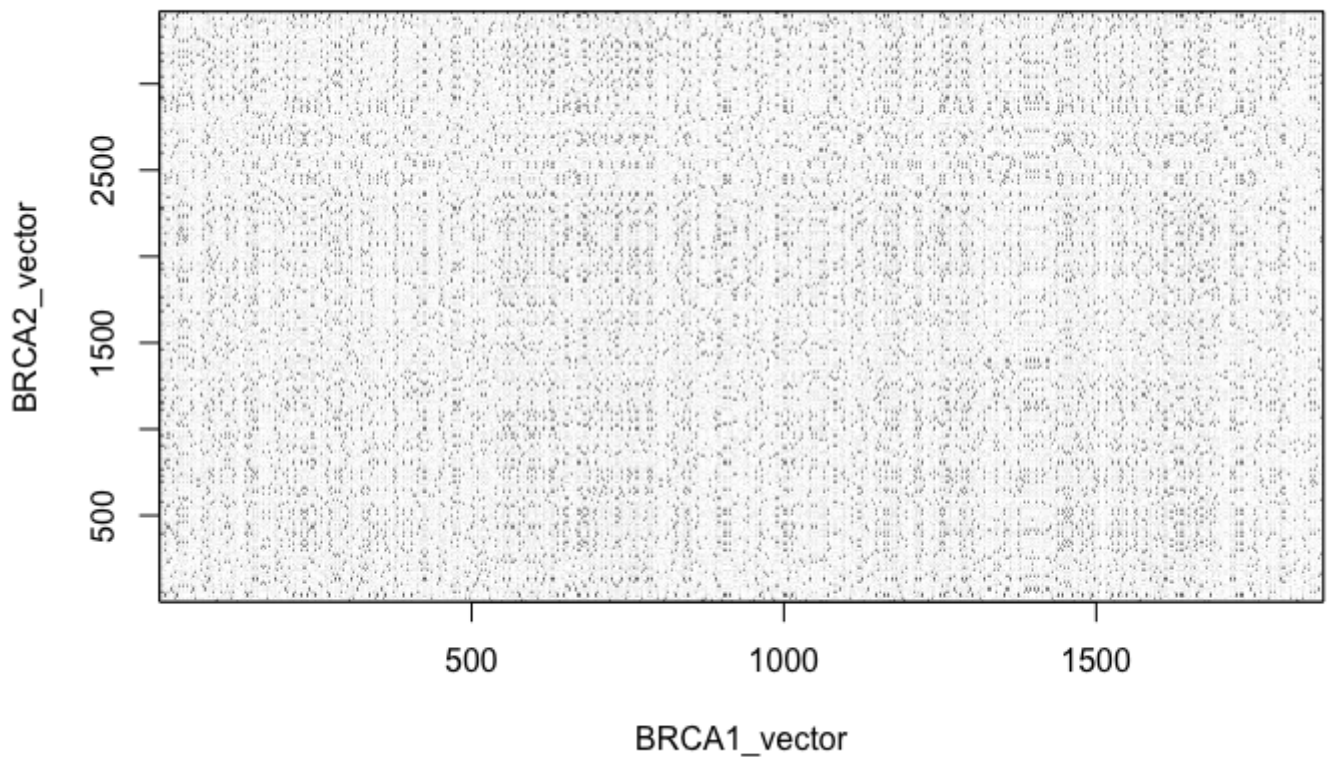
Dotplot: Creating a dotplot is one way to analyze and visualize our results from the pairwise sequence alignment that we performed earlier between the protein sequences of BRCA1 and BRCA2. The BRCA1 sequence is on the x-axis and the BRCA2 sequence is on the y-axis. Similarity between sequences is indicated by a black dot. The dots appear where the sequences match up. Sequences with high similarity would show a distinct line that spans the majority of main diagonal of the dotplot, whereas sequences with low similarity not present a visible diagonal line. (<https://omicstutorials.com/interpreting-dot-plot-bioinformatics-with-an-example/>) (<https://omicstutorials.com/interpreting-dot-plot-bioinformatics-with-an-example/>)

Hide

```
# Convert lists into vectors of single characters
# https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/do.call
# I'm assuming there's an easier way to do this part, I just couldn't figure it out.

do.call("paste", c(brca1_AA, sep = ', ')) -> BRCA1_vector
do.call("paste", c(brca2_AA, sep = ', ')) -> BRCA2_vector

# Create dotplot
BRCA1vs2 <- dotPlot(BRCA1_vector, BRCA2_vector)
```



Hide

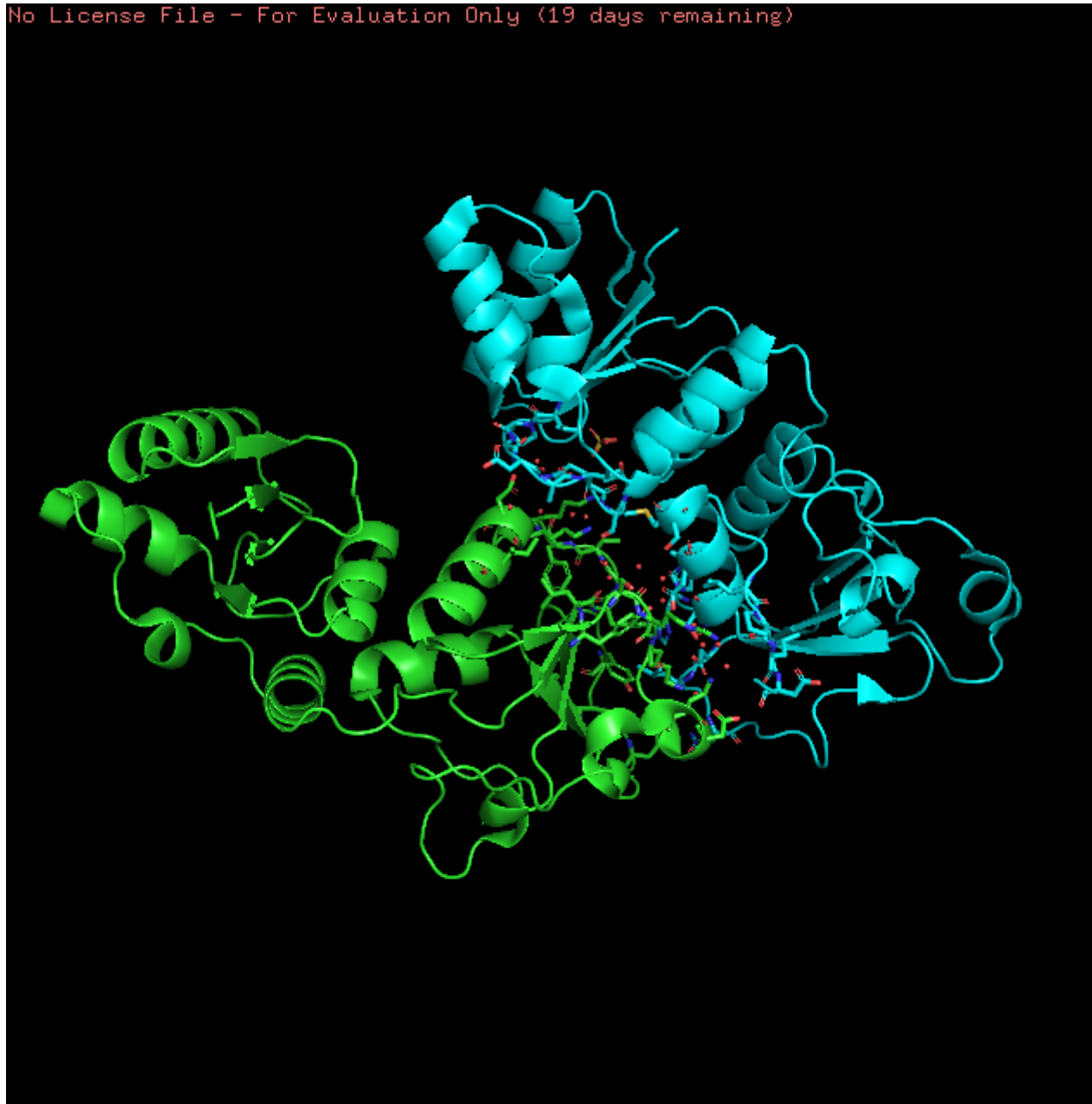
BRCA1vs2

NULL

PyMOL: PyMOL is a molecular visualization system. It can import 3D protein structures straight from the PDB and has functions that allow for analyzing molecules. Pictures and movies can be downloaded from here. It provides many options for different ways to look at molecular structure and even how they align with other molecules. Residues, chains, and atoms can even be labeled and the structure sequences are available to look at as well. (<https://en.wikipedia.org/wiki/PyMOL> (<https://en.wikipedia.org/wiki/PyMOL>))

[Hide](#)

```
# Call knitr package to include images
# https://www.jumpingrivers.com/blog/knitr-include-graphics-external/
library(knitr)
knitr::include_graphics(c("BARD1_BRCT.png", "PALB2_BRCA2.png", "PyMOL align.png"))
```



No License File - For Evaluation Only (19 days remaining)



```
PyMOL>align BARD1_BRCT, PALB2_BRCA2
Match: read scoring matrix.
Match: assigning 430 x 249 pairwise scores.
MatchAlign: aligning residues (430 vs 249)...
MatchAlign: score 27.000
ExecutiveAlign: 149 atoms aligned.
ExecutiveRMS: 2 atoms rejected during cycle 1 (RMSD=8.52).
ExecutiveRMS: 1 atoms rejected during cycle 2 (RMSD=8.32).
ExecutiveRMS: 1 atoms rejected during cycle 3 (RMSD=8.22).
Executive: RMSD = 8.132 (145 to 145 atoms)
```

Hide

```
# Install and call vembedr package to embed video
# install.packages("vembedr")
library("vembedr")
suggest_embed("https://youtu.be/S4B7enDivjM")
```

```
embed_youtube("S4B7enDivjM")  
[Copied to clipboard]
```

[Hide](#)

```
embed_youtube("S4B7enDivjM")
```

[Hide](#)

NA

Analyzing The Results

Pairwise sequence alignment & dotplot:

The alignment score generated from our sequence alignment was -2500.069, indicating that the BRCA1 and BRCA2 protein sequences had very low or no sequence similarity. There is a positive correlation with a higher score and a better alignment. This makes sense though since the BRCA1 protein sequence had nearly double the amount of amino acids than the BRCA2 protein sequence (3418 AA compared to 1863 AA). The dotplot is also very scattered. There are no apparent diagonal lines. There is actually a lot of empty space on the dotplot which makes sense again going back to how the protein sequences differed greatly in amino acid length. The low sequence similarities between BRCA1 and BRCA2 mean that they do have a significant amount of genetic differences. Although their mutations have similar outcomes, the genetic differences between them may be a factor as to why BRCA1 mutations lead to a significantly higher risk of ovarian cancer.

Structural bioinformatics & PyMOL:

From using the structural bioinformatics functions in the Bio3D package, I was able to access information from the PDB files of crystal structures related to BRCA1 and BRCA2, such as the number of models, atoms, chains, and the specific sequences for both structures. The BARD1 BRCT domain is part of BRCA1. BRCA1 interacts with PALB2 which in turn interacts with BRCA2. With a cutoff distance of 50, 17 residues were identified as binding residues between the structures. No residues were identified until the cutoff distance surpassed 42. This means that there are interactions between the BARD1/BRCT complex and the PALB2/BRCA2 complex. Since the cutoff distance is a bit high, this could mean there are other molecules involved in the interactions of these complexes. I

analyzed the same structures in PyMOL by attempting to align them. PyMOL generated an alignment score of 27 between the two structures. These protein complexes match up better than the protein sequences of BRCA1 and BRCA2 and supports the connection between BRCA1 and BRCA2 proteins. In PyMOL, I was able to show the individual structures as well as their alignment. I made sure to include a scene with their protein interfaces as well. The red lines that appeared between the structures are where residues were aligned. There was no evidence of high sequence or structural similarity between BRCA1 and BRCA2, but they do interact together. The protein structural differences in combination with the genetic differences are significant and can be used to explain at least part of the reason why BRCA1 mutations are more likely to cause ovarian cancer.

Code Formatting Requirements

IDENTIFY GLOBAL AND LOCAL VARIABLES + EXPLANATIONS
USE BUILT IN BIOCONDUCTOR FUNCTION +
DESCRIPTION OF FUNCTION INPUT/OUTPUT
HARD-CODING