Building an ESCC database to link clinical properties with molecular mechanistic determinants

(working note)

**Public available databases**

Network of Cancer Gene: <http://ncg.kcl.ac.uk/index.php>

DriverDB: <http://driverdb.ym.edu.tw/DriverDB/intranet/init.do>

KEGG database

DAVID downloadable (EASE): http://david.abcc.ncifcrf.gov/ease/ease1.htm

MsigDB at Broad Institute: <http://www.broadinstitute.org/gsea/msigdb/index.jsp>

**Clinical trial databases:** ([www.clinicaltrials.gov](http://www.clinicaltrials.gov))

As of March 10, 2015, there are 210 ESCC trials at nci website, of which 47 (or of 69 in eastern Asian) are currently conducted in China.

**##===================**

**From the main write up**

**Major milestones in implementation**

Throughout this research process, we plan to implement following parts

1. A web portal for data collection
2. A series of databases
   1. A mutation database
   2. Pathway and mechanistic database
   3. A clinical informatics database
   4. A chemical and drug database
   5. A drug target database
   6. A patient social database
   7. …
3. A front end web-development
4. A database/web-portal paper for NAR in Jan. 2017

**Development staging**

1. Determining the information collection standard/protocol
2. Developing database schema
3. Developing web-portal for data collection
4. Populating the database
5. Developing the web-front end for the database
6. Drafting the NAR paper

**Funding and support**

Incorporate Euclados Bioinformatics Solutions, LLC

One FTE support needed

Getting funded via different routes

1. Determine the information collection protocol
2. A knowledge database to store molecular and clinical information
3. A web development for information retrieval and report, a user level web portal for mechanistic analysis (similar to DAVID, IPA, Jasper etc.)
4. Explore currently existing molecular knowledge database
5. Explore the matured consortium results: 1000 genome, the Broad Institute, TCGA

Public available databases

Network of Cancer Gene: <http://ncg.kcl.ac.uk/index.php>

DriverDB: <http://driverdb.ym.edu.tw/DriverDB/intranet/init.do>

KEGG database

DAVID downloadable (EASE): http://david.abcc.ncifcrf.gov/ease/ease1.htm

MsigDB at Broad Institute: <http://www.broadinstitute.org/gsea/msigdb/index.jsp>

**Deliverable staging**

Determine the information collection standard/protocol

1. Explore currently existing molecular knowledge database
2. Explore the matured consortium results: 1000 genome, the Broad Institute, TCGA

Design a database schema

Develop a web portal for data collection

Build a knowledge database

Develop a web portal for use level’s access

1. A web portal for data collection
2. Design a database schema
3. Determine the information collection protocol
4. A knowledge database to store molecular and clinical information
5. A web development for information retrieval and report, a user level web portal for mechanistic analysis (similar to DAVID, IPA, Jasper etc.)

Reading note (extra information)

As of March 10, 2015, there are 208 ESCC trial at nci website, of which 47 are currently conducted in China

Sox2 tends to interact with p63 as opposed to Oct4 in embryonic stem cells