Building an ESCC database to link clinical property with molecular mechanistic determinant

**Introduction**

In order to facilitate biomedical analysis of esophagus squamous cell carcinoma (ESCC), we are building a knowledgebase that links clinical properties including the diagnosis end points to the known molecular information, i.e. genomic variants (snv, indel, and other genomic abnormality) to render pathway and network information for biomedical research and provide basis for diagnosis and relapse prediction model

**Purpose to serve the following aspects and/or meeting the unmet need**

1. Driver mutation vs. passenger mutation
2. In correct database information (sox2 --> p60 vs. etc) in the existing pathway analysis databases (i.e.
3. Mutation bank, i.e. dbSNP etc.
4. Interaction with drug target
5. Collection of current clinical trials associated with ESCC, possible intervention links to “cancer medicine”, which shall be linked somehow to “chemical compound database” (if any)

**Main implementation**

Throughout this research process, we plan to implement following parts

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.)

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>

**Development 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