MOLECULAR ONCOLOGY 7 (2013) 791–798

available at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/molonc

Tumor Suppressors Status in Cancer Cell Line Encyclopedia

Dmitriy Sonkin[a], *Mehedi Hassan*[a], *Denis J. Murphy*[a], *Tatiana V. Tatarinova*[a,b,*]

[a]*University of South Wales, Pontypridd CF37 1DL, Wales, UK*
[b]*Laboratory of Applied Pharmacokinetics, University of Southern California, USA*

ARTICLE INFO

Article history:
Received 25 January 2013
Received in revised form
17 March 2013
Accepted 2 April 2013
Available online 11 April 2013

Keywords:
Tumor suppressor
Cancer cell line
CCLE
Loss of function
DNA methylation
Epigenetics

ABSTRACT

Tumor suppressors play a major role in the etiology of human cancer, and typically achieve a tumor-promoting effect upon complete functional inactivation. Bi-allelic inactivation of tumor suppressors may occur through genetic mechanisms (such as loss of function mutation, copy number (CN) loss, or loss of heterozygosity (LOH)), epigenetic mechanisms (such as promoter methylation or histone modification), or a combination of the two. We report systematically derived status of 69 known or putative tumor suppressors, across 799 samples of the Cancer Cell Line Encyclopedia. In order to generate such resource we constructed a novel comprehensive computational framework for the assessment of tumor suppressor functional "status". This approach utilizes several orthogonal genomic data types, including mutation data, copy number, LOH and expression. Through correlation with additional data types (compound sensitivity and gene set activity) we show that this integrative method provides a more accurate assessment of tumor suppressor status than can be inferred by expression, copy number, or mutation alone. This approach has the potential for a more realistic assessment of tumor suppressor genes for both basic and translational oncology research.

1. Introduction

Tumor suppressor genes encode proteins that normally inhibit tumor formation caused by abnormal cellular proliferation. Tumor suppressor proteins can participate in a variety of processes such as negative regulation of the cell cycle, positive regulation of apoptosis, regulation of DNA damage response, or other mechanisms (Stanbridge, 1990). The list of tumor suppressor genes includes such names as TP53 (tumor protein p53), RB1 (retinoblastoma), APC (adenomatous polyposis coli), and BRCA1 (breast cancer 1, early onset). The inactivation of these and other tumor suppressor genes plays a major role in many types of cancer (Jones and Thompson, 2009).

Unlike proto-oncogenes, where a single mutation can be dominant and lead to cellular transformation, a single mutation in a tumor suppressor gene is normally recessive as long as there is a second functional copy of the gene (Knudson, 1971). However, loss of function of both tumor suppressor alleles may promote tumor growth or survival providing that the loss of function is nearly or totally complete. It is possible to infer loss of function of tumor suppressor genes through a number of genomic measurements, such as transcript expression, DNA copy number, and mutation.

The Cancer Cell Line Encyclopedia (CCLE, http://www.broadinstitute.org/ccle/home) is a recently compiled public resource that contains gene expression, chromosomal copy number and massively parallel sequencing data from nearly 1000 cancer cell lines (Barretina et al., 2012). These matched datasets allow for the examination of distinct mechanisms of tumor suppressor inactivation and also for

* *Corresponding author*. University of South Wales, Pontypridd CF37 1DL, Wales, UK.
E-mail addresses: dmitriy.sonkin@southwales.ac.uk (D. Sonkin), tatiana.tatarinova@lapk.org (T.V. Tatarinova).
1574-7891/$ – see front matter
http://dx.doi.org/10.1016/j.molonc.2013.04.001

"integrative analyses" of orthogonal data types. In order to further extend the utility of the CCLE both for basic and translational oncology research communities, we have built a comprehensive computational framework for assessing the functional status of tumor suppressor genes. We have then applied this framework to 69 known or putative tumor suppressors across the CCLE. In this paper, we demonstrate that this integrative method provides a more powerful and more reliable tool for tumor suppressor gene analysis than simply utilizing individual datasets.

2. Materials and methods

We compiled a list of 82 well-known and putative tumor suppressor genes. Among these, 69 genes have mutation, copy number and expression data available and, therefore, were used for the present analysis (Supp. Table 1). We assembled information from the literature on known loss of function missense mutations (Table 1). At this time the number of clearly validated loss of function missense mutations is small (only 38 entries covering 7 genes). However, it is likely that there are other *bona fide* losses of function missense mutations that have not been sufficiently validated or annotated.

Affymetrix U133Plus2 mRNA expression, Affymetrix SNP 6.0 data, OncoMap mutation calls (MacConaill et al., 2009), exome data sequencing (Hodges et al., 2007), and pharmacological profiling data are available at the CCLE website. All expression values are MAS5 normalized, with a 2% trimmed mean of 150 (Hubbell et al., 2002). We summarized cutoffs used for expression, copy number, and mutation data in Table 2.

We have divided mechanisms of inactivation of tumor suppressors into three categories. Figure 1 illustrates each subcategory with a simplified diagram.

The first category "G" is based completely on genetic mechanisms of inactivation of both alleles (Stanbridge, 1990; Ponder, 2001) and, therefore, can be considered as the highest confidence category.

The genetic category can be subdivided further into 2 subcategories:

1. The sub-category "G-M" is based on a homozygous nonsense, frame shift, loss of function missense mutation or heterozygous/homozygous dominant negative mutation.
2. The sub-category "G-D" is based on deletion of both alleles (bi-allelic loss).

One way for a gene to appear in the sub-category "G-M" is to have LOH status derived from Affymetrix SNP 6.0 data and a homozygous mutation deduced from the exome sequencing data. Any nonsense or frame shift mutation is considered to lead to loss of function; however, only validated loss of function missense mutations from Table 1 are used. Figure 1 illustrates a sub-category "G-M" with the most likely scenario being the

Table 1 – Known loss of function missense mutations.

Gene	ENTREZ_ID	AA.Change	dbSNP	Dominant negative
CDKN2A	1029	H83Y		
CDKN2A	1029	D84Y	rs11552822	
CDKN2A	1029	D108Y		
CDKN2A	1029	P114L		
MLH1	4292	V384D		
PTEN	5728	R130G		
PTEN	5728	R130Q		
PTEN	5728	R173C		
PTEN	5728	R173H		
RB1	5925	C706F		
STK11	6794	D194N		
STK11	6794	D194V		
STK11	6794	E199K		
STK11	6794	P281L		
TP53	7157	V143A		N
TP53	7157	V157F		Y
TP53	7157	R158L		Y
TP53	7157	R158H		N
TP53	7157	R175H	rs28934578	Y
TP53	7157	Y220C		Y
TP53	7157	M237I		N
TP53	7157	G245S	rs28934575	Y
TP53	7157	R248Q	rs11540652	Y
TP53	7157	R248W		Y
TP53	7157	R249S		Y
TP53	7157	R273C		Y
TP53	7157	R273H	rs28934576	Y
TP53	7157	R273L		Y
TP53	7157	R280K		N
TP53	7157	R280S		N
TP53	7157	R280T		N
TP53	7157	R282G		N
TP53	7157	R282W		Y
VHL	7428	P81S	rs5030806	
VHL	7428	L85P	rs5030828	
VHL	7428	L89H	rs5030807	
VHL	7428	L158Q		
VHL	7428	R167W	rs5030820	

Table 2 – Cutoffs for expression, copy number, and mutation data.

Copy number (CN) ratio <0.6 indicates "allelic loss".
CN ratio is the ratio of signal intensity in a tumor sample versus normal reference samples normalized to total DNA quantity; thus a CN ratio of 1 corresponds to a diploid locus.
Copy number ratio <0.25 indicates "bi-allelic loss", or complete loss.
Copy number ratio >0.9 indicates the presence of both alleles.
Gene expression <32 is considered to be "not expressed", when the mean and median expression of this gene across all cell lines are above 100. For calculation of mean and median gene expression values, we discarded cell lines with CN ratio below 0.25, in order to decrease artificial under-estimation of expression distributions of cell lines with remaining functional DNA.
Gene expression >300 is considered as a "high confidence" level of expression.
Mutation data: a minimum of 20 mutant reads defines a "trusted mutation", this is a conservative cutoff designed to minimize false positive calls. No more than one read for the wild type allele is allowed for homozygous calls.
(For reference, hybrid capture exome sequencing was performed to an average depth of 60-fold.)
OncoMap mutations are considered to be heterozygous.