TCGA Data License

TCGA (NIH/NCI) <tcga@mail.nih.gov>   Mon, Jul 18, 2016 at 11:33 AM
To: Daniel Himmelstein <daniel.himmelstein@gmail.com>, "TCGA (NIH/NCI)" <tcga@mail.nih.gov>
Cc: Casey Greene <csgreene@mail.med.upenn.edu>, Gregory Way <gregway@mail.med.upenn.edu>

Hi Daniel,

The TCGA open access data is in the public domain. You may redistribute as you see fit. We only ask that you attribute the data to TCGA.

Best,

Amy Blum
Communications Manager
NCI | CCG | TCGA

From: Daniel Himmelstein [mailto:daniel.himmelstein@gmail.com]
Sent: Saturday, July 16, 2016 6:51 PM
To: TCGA (NIH/NCI)
Cc: Casey Greene; Gregory Way
Subject: TCGA Data License

Greetings,

I'm helping lead a collaborative open source project, called Cognoma, which helps biologists perform machine learning on TCGA data. Our use of TCGA data will exclusively be of content in the Open Access Data Tier. I couldn't find any information on the license of TCGA data. I did see the Publication Guidelines, but it was unclear what reuse rights, if any, those guidelines grant.

We would like to redistribute TCGA datasets and publish derivatives under an open license, preferably CC0. Is TCGA data in the public domain as work prepared by employees of the United States government? If not and copyright protection does apply to TCGA data, is there a license granting rights to the public? If the data is subject to copyright and not publicly licensed, can Project Cognoma receive permission to distribute the data under an open license such as CC0? These questions are all in the context of the Open Access data tier, not the Controlled Access data tier.

Best,
Daniel