Cancer Variant Intersection Analysis: Additional Datasets

Lucas Lochovsky
Annotation subgroup
Gerstein lab, Yale
Sept 19, 2012

Data

- Prostate Cancer
  - Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* (2011).
  - Seven samples
  - ~29,000 variants
  - ~0.7% exome
- Melanoma
  - Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2009).
  - One sample
  - ~33,000 variants
  - ~0.9% exome

Data

- Breast Cancer
  - Nik-Zainal, S. *et al.* Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell* **149**, 979–993 (2012).
  - 21 samples
  - ~180,000 variants
  - ~0.8% exome
  - Looking at “gene_type” column, it appears that variants are ~30% protein coding
  - But “mut_type” column indicates that most of those are intron variants

Methods

- Separate driver and passenger mutations
  - Using CHASM (**C**ancer-Specific **H**igh-throughput **A**nnotation of **S**omatic **M**utations)
  - Wong, W. C. *et al.* CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. *Bioinformatics* **27**, 2147–2148 (2011).
- Investigate enrichments/depletions of drivers and passengers in phase 1 1KG coding variants through intersection analysis
  - Common vs. rare 1KG variants
  - Nonsynonymous vs. synonymous 1KG variants
- Random expectation: Average of 10,000 runs using randomized cancer variant coordinates

Significant Results: Berger Prostate Cancer

Drivers

- Only one driver variant overlapped 1KG variant
  - Rare, nonsynonymous

Passengers

Comparison	Actual:Random Percentage Ratio	**p-value**	**95% CI (random dist)**
Passengers vs. all 1KG coding	2.38	8.39E-04	[0.63018742, 9.47501258]
Passengers vs. 1KG common coding	4.43	0.000160	[-1.021558522, 3.280358522]
Passengers vs. 1KG rare coding	1.90	0.0398	[-0.10245336, 7.46765336]
Passengers vs. 1KG nonsyn coding	3.04	1.82E-04	[-0.43441278, 6.34621278]
Passengers vs. 1KG syn coding	1.39	0.280	[-0.74319146, 5.05239146]

Significant Results: Pleasance Melanoma

Drivers

Comparison	Actual:Random Percentage Ratio	p-value	95% CI (random dist)
Drivers vs. all 1KG coding	2.38	0.0347	[-0.872331252, 4.234531252]
Drivers vs. 1KG common coding	5.46	0.00344	[-0.842927884, 1.575327884]
Drivers vs. 1KG rare coding	1.62	0.243	[-0.949376666, 3.423976666]
Drivers vs. 1KG nonsyn coding	1.02	0.490	[-0.97056178, 2.92396178]
Drivers vs. 1KG syn coding	4.13	0.00360	[-0.966859322, 2.418059322]

Passengers

Comparison	Actual:Random Percentage Ratio	p-value	95% CI (random dist)
Passengers vs. all 1KG coding	1.40	0.167	[0.97642116, 10.45117884]
Passengers vs. 1KG common coding	1.56	0.261	[-0.95666989, 3.52406989]
Passengers vs. 1KG rare coding	1.44	0.185	[0.08550592, 8.24829408]
Passengers vs. 1KG nonsyn coding	1.50	0.179	[-0.30308568, 6.96128568]
Passengers vs. 1KG syn coding	1.22	0.364	[-0.67620584, 5.58520584]

Significant Results: Nik-Zainal Breast Cancer

Drivers

- There were zero intersections between drivers and 1KG variants

Passengers

Comparison	Actual:Random Percentage Ratio	p-value	95% CI (random dist)
Passengers vs. all 1KG coding	1.53	4.73E-04	[25.46989484, 49.23870516]
Passengers vs. 1KG common coding	2.03	0.001373404	[2.6067239, 14.1334761]
Passengers vs. 1KG rare coding	1.43	0.0104	[17.09457002, 37.41062998]
Passengers vs. 1KG nonsyn coding	1.92	5.50E-06	[12.68312814, 31.01627186]
Passengers vs. 1KG syn coding	1.19	0.220	[8.06816372, 23.83483628]

Caveat

- Sample size *n* for the Halaban data is 275, but the highest *n* in these datasets is 21

For the Future…

- Intersection analysis with TCGA cancer data is ongoing