

## TCGA Mutation/Variation Calling Benchmark 4 at CGHub

The Mutation Calling Benchmark 4 dataset consists of whole genome cell line data. This mutation calling experiment is open to groups outside TCGA. We encourage participation from all groups with an interest in calling mutations in the paired tumor/normal context of cancer genomics. The overall purpose of this benchmark exercise is comparative evaluation of somatic mutation calls on single nucleotide variants (SNVs) and structural variants (SVs) under a variety of conditions designed to simulate the effects of tumor purity (i.e. normal contamination) and subclonal expansions in a controlled way. In addition, we can eventually know the veracity of the called mutations because the cell lines are publicly available, making all mutation calls amenable to experimental validation.

The BAMs for TCGA Benchmark 4 are available from CGHub (listed below).

NOTE: these have been updated since originally published, the 10 which had artificially spiked variants ("spiked1" in their filenames) have been regenerated due to an issue with their copy number in the prior versions.

**\*\*ONLY FOR USE IN BENCHMARK EXERCISE(s), these are simulated mixtures of cell line derived DNA, they are NOT real natural/medical phenomena\*\*** For further details see: [Benchmark 4 Instructions PDF](#) and/or email Adam Ewing at: [ewingad@soe.ucsc.edu](mailto:ewingad@soe.ucsc.edu).

To download these do the following:

- **Download GeneTorrent**

Please see the CGHub User's Guide to install gtdownload.

- **Download (right-click and select "Save Link As...") public access token file**
- **Run the following:**

To get the full manifest of the TCGA Benchmark 4 BAM files, run this using cgquery (also available from the same page as GeneTorrent above):

```
cgquery -o b4.xml "study=TCGA_MUT_BENCHMARK_4&state=live"
```

you can then either download them all at once (if you have the space and time):

```
gtdownload -c https://cghub.ucsc.edu/software/downloads/cghub_public.key -vv -d b4.xml
```

Or if you prefer to do them one at a time or custom scripted:

```
gtdownload -c https://cghub.ucsc.edu/software/downloads/cghub_public.key -vv -d  
{uuid_from_list_below}
```

Note when running gtdownload: If you were not able to install in /usr/bin (due to not having admin rights for instance), run this instead, substituting your home directory (or whatever is the full path to where gtdownload was untarred) for the first two "/path/to"s:

```
~/cghub/bin/gtdownload -C ~/cghub/share/GeneTorrent -c  
https://cghub.ucsc.edu/software/downloads/cghub_public.key -vv -d  
{uuid_from_list_below}
```

uuid	description	filename	Size (Giga bytes)
ad3d4757-f358-40a3-9d92-7 42463a95e88	TCGA BENCHMARK CELL LINE: HCC1143 TUMOR 50x	G15511.HCC1143.1.ba m	255.8
f0eaa94b-f622-49b9-8eac-e 4eac6762598	TCGA BENCHMARK CELL LINE: HCC1143 NORMAL 60x	G15511.HCC1143_BL.1 .bam	305.8
4bcf3463-ea9c-414e-a1f5-9 48f72477602	TCGA BENCHMARK CELL LINE: HCC1143 NORMAL 30x (use for comparisons)	HCC1143.NORMAL.30x .compare.bam	143.5
6d8044f7-3f63-487c-9191-a dfeed4e74d3	TCGA BENCHMARK CELL LINE: HCC1954 TUMOR 58x	G15512.HCC1954.1.ba m	279.2
34c9ff85-c2f8-45dc-b4aa-fba 05748e355	TCGA BENCHMARK CELL LINE: HCC1954 NORMAL 71x	G15512.HCC1954_BL.1 .bam	338.5

99d19ebd-832c-4f2d-b97f-51743a2cf9a2	TCGA BENCHMARK CELL LINE: HCC1954 NORMAL 30x (use for comparisons)	HCC1954.NORMAL.30x .compare.bam	137.2
5462741b-774b-41cc-b3a2-d3cc7eaa676	UCSC ARTIFICIAL MIXED SAMPLE: 5% HCC1143BL 95% HCC1143	HCC1143.mix1.n5t95.ba m	136.9
02d8b3de-b043-4bfa-9130-adc18195313f	UCSC ARTIFICIAL MIXED SAMPLE: 40% HCC1954BL 60% HCC1954	HCC1954.mix1.n40t60.b am	133.0
47e16d4c-fe0f-4ce0-9678-645efe53ca30	UCSC ARTIFICIAL MIXED SAMPLE: 60% HCC1143BL 40% HCC1143	HCC1143.mix1.n60t40.b am	137.1
712a71eb-e62d-46e4-acd6-883b4dbc5053	UCSC ARTIFICIAL MIXED SAMPLE: 20% HCC1143BL 80% HCC1143	HCC1143.mix1.n20t80.b am	137.3
7943cc03-39ac-4bff-9f40-d35a782400eb	UCSC ARTIFICIAL MIXED SAMPLE: 60% HCC1954BL 40% HCC1954	HCC1954.mix1.n60t40.b am	132.6
e90b4f61-6ac1-4632-aea0-051438e9ab22	UCSC ARTIFICIAL MIXED SAMPLE: 80% HCC1143BL 20% HCC1143	HCC1143.mix1.n80t20.b am	136.5
360b4736-6c5e-48df-af58-c1cf51609350	UCSC ARTIFICIAL MIXED SAMPLE: 80% HCC1954BL 20% HCC1954	HCC1954.mix1.n80t20.b am	131.6

97d04e21-9ec1-4554-9a60-f 14b2974a8a7	UCSC ARTIFICIAL MIXED SAMPLE: 20% HCC1954BL 80% HCC1954	HCC1954.mix1.n20t80.b am	132.8
708f1069-f91d-4111-a3e9-c bc690b9dda6	UCSC ARTIFICIAL MIXED SAMPLE: 40% HCC1143BL 60% HCC1143	HCC1143.mix1.n40t60.b am	137.4
048de7d8-682f-4e72-99d9-b e2d072af67b	UCSC ARTIFICIAL MIXED SAMPLE: 95% HCC1143BL 5% HCC1143	HCC1143.mix1.n95t5.ba m	135.7
54bc737c-ecd9-4dea-8b44-8 c2197368da8	UCSC ARTIFICIAL MIXED SAMPLE: 5% HCC1954BL 95% HCC1954	HCC1954.mix1.n5t95.ba m	132.0
36e0c1a5-9423-4887-90e7- 5ce7d031fe9f	UCSC ARTIFICIAL MIXED SAMPLE: 95% HCC1954BL 5% HCC1954	HCC1954.mix1.n95t5.ba m	130.1
111f6781-76a7-444b-a7b8-b 2ef102fd95f	UCSC Mixed spike-in sample for TCGA Benchmark 4: 25% HCC1954BL 65% HCC1954 10% HCC1954 simulated subclone	HCC1954.spiked1.n25t6 5s10.bam	123.3
26a7c278-6e69-4138-a9d9- c8f0abb53a1c	UCSC Mixed spike-in sample for TCGA Benchmark 4: 25% HCC1954BL 70% HCC1954 5% HCC1954 simulated subclone	HCC1954.spiked1.n25t7 0s5.bam	130.4

946478e8-434b-4310-af2c-c 1604749cc10	UCSC Mixed spike-in sample for TCGA Benchmark 4: 25% HCC1954BL 35% HCC1954 40% HCC1954 simulated subclone	HCC1954.spiked1.n25t3 5s40.bam	119.0
946a6992-396c-4160-9fb6-6 3093f09af4e	UCSC Mixed spike-in sample for TCGA Benchmark 4: 25% HCC1954BL 74% HCC1954 1% HCC1954 simulated subclone	HCC1954.spiked1.n25t7 4s1.bam	132.9
9a059503-90d7-4ae6-bd75- 824a90ac3088	UCSC Mixed spike-in sample for TCGA Benchmark 4: 25% HCC1143BL 74% HCC1143 1% HCC1143 simulated subclone	HCC1143.spiked1.n25t7 4s1.bam	136.9
9bfdf4db-4e91-4dae-a299-8f b9f5cc7aa8	UCSC Mixed spike-in sample for TCGA Benchmark 4: 25% HCC1143BL 70% HCC1143 5% HCC1143 simulated subclone	HCC1143.spiked1.n25t7 0s5.bam	134.4
9e8465eb-c99e-4f20-a358-9 7513c17ebfe	UCSC Mixed spike-in sample for TCGA Benchmark 4: 25% HCC1143BL 65% HCC1143 10% HCC1143 simulated subclone	HCC1143.spiked1.n25t6 5s10.bam	127.2
c193bb0b-3a9c-4f7d-a6d0-f 6a979a9e5f4	UCSC Mixed spike-in sample for TCGA	HCC1954.spiked1.n25t5 5s20.bam	120.1

	Benchmark 4: 25% HCC1954BL 55% HCC1954 20% HCC1954 simulated subclone		
d855c8c8-6be8-44b5-8bfc-1 a3642be4022	UCSC Mixed spike-in sample for TCGA Benchmark 4: 25% HCC1143BL 35% HCC1143 40% HCC1143 simulated subclone	HCC1143.spiked1.n25t3 5s40.bam	122.8
ecf60c24-a775-4e5b-b5c9-7 c9c87b47994	UCSC Mixed spike-in sample for TCGA Benchmark 4: 25% HCC1143BL 55% HCC1143 20% HCC1143 simulated subclone	HCC1143.spiked1.n25t5 5s20.bam	123.9
54b4c169-bdef-4d71-99be-f 6458d753f1c	UCSC Mixed low coverage spike-in sample for TCGA Benchmark 4: 25% HCC1143BL 65% HCC1143 10% HCC1143 simulated subclone **ONLY FOR USE IN BENCHMARK EXERCISE**	HCC1143.7x.n25t65s10. bam	33.0
5d563fa8-33f1-4959-be0b-b 1c5268bca99	High-cover sample for TCGA Benchmark 4: HCC1954 NORMAL 7x (use for low coverage comparisons) **ONLY	HCC1954.NORMAL.7x. compare.bam	32.0

	FOR USE IN BENCHMARK EXERCISE**		
6344cbbb-aed9-4173-9450- 2507a55f4c9f	High-cover sample for TCGA Benchmark 4: HCC1143 NORMAL 7x (use for low coverage comparisons) **ONLY FOR USE IN BENCHMARK EXERCISE**	HCC1143.NORMAL.7x. compare.bam	33.2
6b3b437e-24bc-45fd-8363-6 38afe641aca	UCSC Mixed spike-in low coverage sample for TCGA Benchmark 4: 25% HCC1954BL 65% HCC1954 10% HCC1954 simulated subclone **ONLY FOR USE IN BENCHMARK EXERCISE**	HCC1954.7x.n25t65s10. bam	32.1
cd768f1a-10dc-4d3e-b119-4 7a70ed22335	UCSC Mixed low coverage spike-in sample for TCGA Benchmark 4: 25% HCC1143BL 55% HCC1143 20% HCC1143 simulated subclone **ONLY FOR USE IN BENCHMARK EXERCISE**	HCC1143.7x.n25t55s20. bam	32.7
eea56631-c802-47ff-8118-3 ed40d10302b	UCSC Mixed spike-in low coverage sample for TCGA Benchmark 4: 25% HCC1954BL 55%	HCC1954.7x.n25t55s20. bam	31.9

	HCC1954 20% HCC1954 simulated subclone **ONLY FOR USE IN BENCHMARK EXERCISE**		
--	---	--	--

updated 2013-01-03T16:56

created 2012-08-20T17:00