TP53 Allelic State in MSK-Impact Cohort

03/30/2020

Outline

Copy Number Vocabulary and Abbreviations (1/3)

Main Copy-Number Events

- <u>LOSS</u>
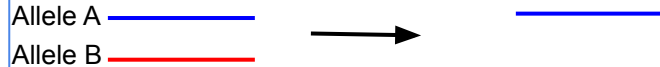
  Allele A —————— → ——————

  Allele B ——————

- <u>Copy-Neutral Loss of Heterozygosity (CNLOH)</u>

  —————— → ——————

  ——————   ——————

- <u>GAIN</u>

  —————— → ——————

  ——————   ——————

          ——————
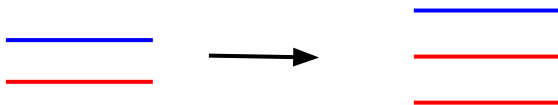
- <u>Whole-Genome Doubling (WGD)</u>

  ——————

  ——————

  ↓

  ——————

  ——————

  ——————

  ——————
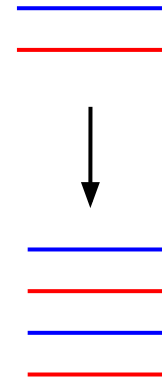
Copy Number Vocabulary and Abbreviations (2/3)

- **Total Copy Number (tcn):**
  → Number of alleles for a given locus

- **Major Copy Number (mcn):**
  → The most-represented allele in the cell

  $$tcn = lcn + mcn$$

- **Lower Copy Number (lcn):**
  → The less-represented allele in the cell

- <u>Diploid</u>

  tcn = 2 ; mcn = 1 ; lcn = 1

- <u>LOSS</u>

  tcn = 1 ; mcn = 1 ; lcn = 0

- <u>CNLOH</u>

  tcn = 2 ; mcn = 2 ; lcn = 0
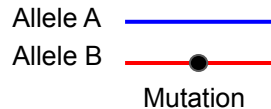
- <u>Tetraploid</u>

  tcn = 4 ; mcn = 2 ; lcn = 2

Copy Number Vocabulary and Abbreviations (3/3)

Mono-Allelic State

At least one wild-type allele in the cells.

Examples

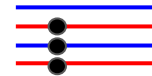- Diploid

  Allele A ⸻
  Allele B ⸺●⸺
  Mutation

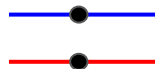- Loss without mutation

  ⸻

- Tetraploid with 3 mutations

Bi-Allelic State

All the alleles of the gene are hit, by mutations or copy number events such as Loss or cnLOH.

→ For a diploid karyotype, all alleles = 2 alleles

Examples

- Diploid with 2 mutations on both alleles

- Loss and 1 mutation

- 1 mutation, then cnloh
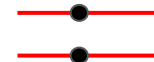
Motivation

TP53 Allelic State in MDS → Pan Cancer Analysis

TP53 Allelic State in MDS

- TP53 Allelic State has important consequences on clinical outcomes
    - Higher risk of death for bi-allelic hits compared to mono-allelic
    - Bi-allelic hits seem to be a driver of disease progression (MDS → AML)

Pan Cancer Analysis: MSK-Impact Cohort

- What are the representations of TP53 allelic states across cancer types?
- Is there significant clinical / phenotypic differences between mono-allelic and bi-allelic states ?
    - What are the cancers where the differences are most significant? The one where there are no differences?

MSK-Impact Cohort

Description of the data

Four different files

1. **Cohort file**

- Entire cohort samples
- Paths to raw data
- Run versions and parameters
- Purity / Ploidy

2. **TP53 mutation file**

- Only TP53 mutant samples
- Mutation Information
  - *Start, End,*
  - *Protein Change, Mutation Type, Ref Allele, Tumor Allele*
  - *mutation_effect, oncogenic,*
  - *vaf, ccf*

Description of the data

Four different files

1. **Cohort file**

- Entire cohort samples
- Paths to raw data
- Run versions and parameters
- Purity / Ploidy

2. **TP53 file**

- Only TP53 mutant samples
- Mutation Information
    - *Start, End,*
    - *Protein Change, Mutation Type, Ref Allele, Tumor Allele*
    - *mutation_effect, oncogenic,*
    - *vaf, ccf*

3. **TP53 Copy-number File**

- Entire cohort samples (dif: 446 samples)
- total copy number, lower copy number
- copy number state

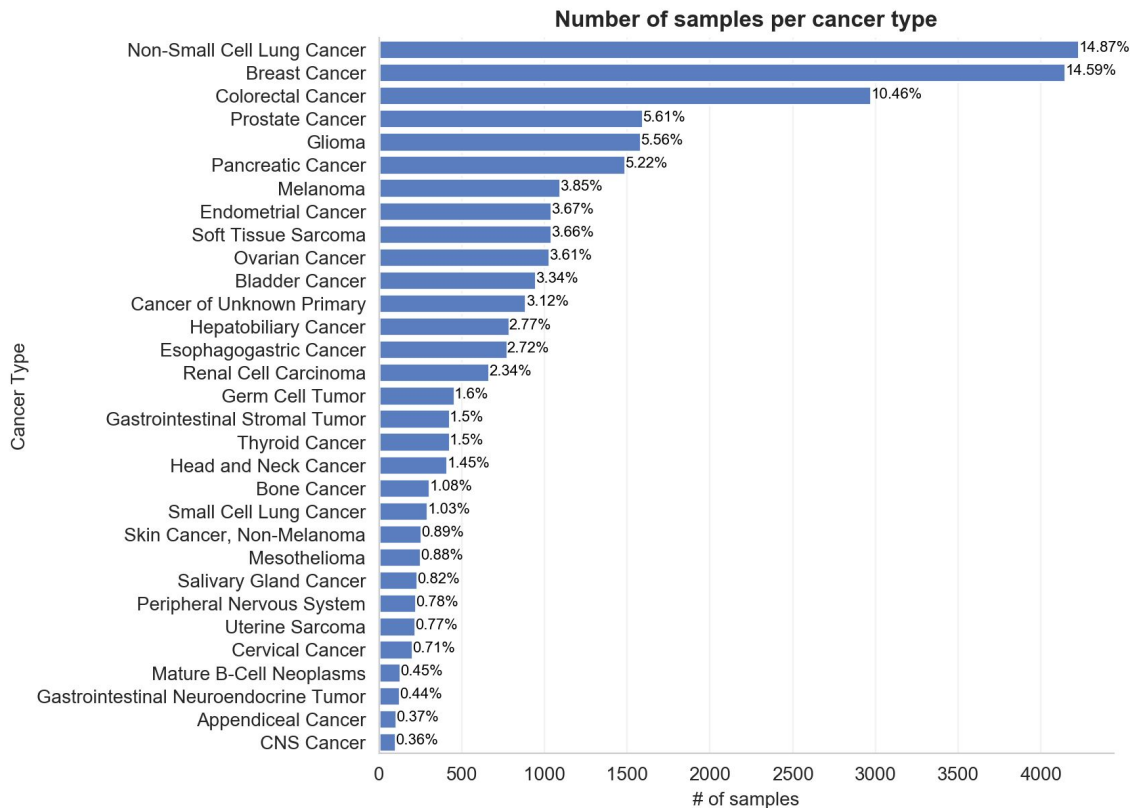4. **Arm Level File**

- Entire cohort samples (dif: 439 samples)
    - All arms (1p to 23q)
    - threshold: frac_of_arm = 0.8
- arm, tcn ,lcn, cn_state
- an All-Chromosome event (loss, cnloh ...) is encoded as two lines with same cn state and frac=1
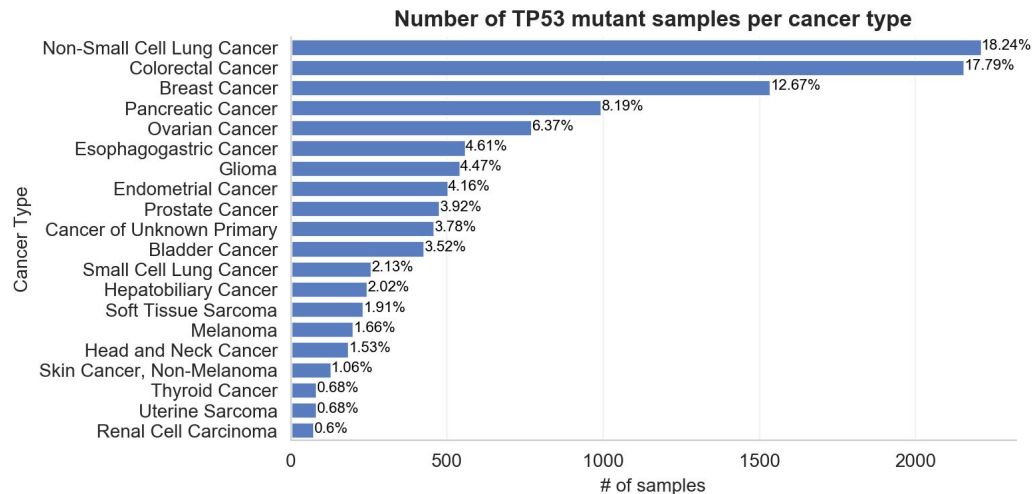
Overall Numbers (1/2)

Entire Cohort

- Number of samples in the cohort: **29 259**
- Number of patients in the cohort: **27 021**
- Proportion of Metastatic Samples: **41.3%**
- **66** Tumor types

Number of samples per cancer type

Cancer Type	%
Non-Small Cell Lung Cancer	14.87%
Breast Cancer	14.59%
Colorectal Cancer	10.46%
Prostate Cancer	5.61%
Glioma	5.56%
Pancreatic Cancer	5.22%
Melanoma	3.85%
Endometrial Cancer	3.67%
Soft Tissue Sarcoma	3.66%
Ovarian Cancer	3.61%
Bladder Cancer	3.34%
Cancer of Unknown Primary	3.12%
Hepatobiliary Cancer	2.77%
Esophagogastric Cancer	2.72%
Renal Cell Carcinoma	2.34%
Germ Cell Tumor	1.6%
Gastrointestinal Stromal Tumor	1.5%
Thyroid Cancer	1.5%
Head and Neck Cancer	1.45%
Bone Cancer	1.08%
Small Cell Lung Cancer	1.03%
Skin Cancer, Non-Melanoma	0.89%
Mesothelioma	0.88%
Salivary Gland Cancer	0.82%
Peripheral Nervous System	0.78%
Uterine Sarcoma	0.77%
Cervical Cancer	0.71%
Mature B-Cell Neoplasms	0.45%
Gastrointestinal Neuroendocrine Tumor	0.44%
Appendiceal Cancer	0.37%
CNS Cancer	0.36%

TP53 mutant cohort

- Number of TP53 mutant samples: **12 731 (43.5%)**
- Number of TP53 mutant patients: **11 982**
- Proportion of Metastatic Samples: 43.2%
- 53 tumor types

Number of TP53 mutant samples per cancer type

Cancer Type	Percentage
Non-Small Cell Lung Cancer	18.24%
Colorectal Cancer	17.79%
Breast Cancer	12.67%
Pancreatic Cancer	8.19%
Ovarian Cancer	6.37%
Esophagogastric Cancer	4.61%
Glioma	4.47%
Endometrial Cancer	4.16%
Prostate Cancer	3.92%
Cancer of Unknown Primary	3.78%
Bladder Cancer	3.52%
Small Cell Lung Cancer	2.13%
Hepatobiliary Cancer	2.02%
Soft Tissue Sarcoma	1.91%
Melanoma	1.66%
Head and Neck Cancer	1.53%
Skin Cancer, Non-Melanoma	1.06%
Thyroid Cancer	0.68%
Uterine Sarcoma	0.68%
Renal Cell Carcinoma	0.6%

# of samples

Data Structuration (1/2)

Data Structuration (2/2)

Clinical Data
- Cancer Type
- Sample Type
- Survival Information
- Purity
- Ploidy
- Samples per patient

TP53 mutations
- # of mutations
- VAF | max_VAF | CFF | Mutation Type
- Hotspot

TP53 locus Copy-Number
- Total Copy Number (tcn)
- Copy Number State
- Expected number of mutant copies
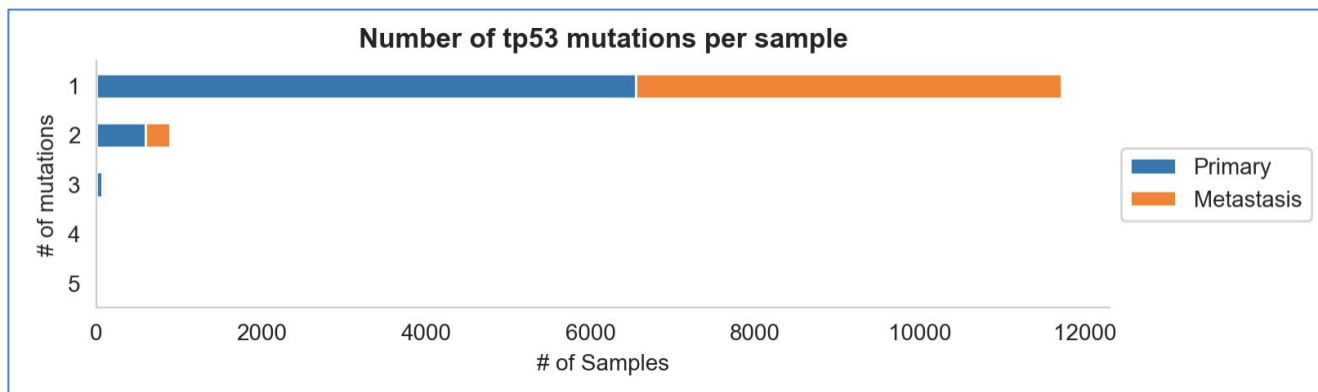- Major Copy Number (mcn)
- Lower Copy Number (lcn)

Genome-wide instability
- WGD
- # of chr with loss, gain, cnloh [arm size level]
- fraction of genome altered

Tumor burden
- TMB | MSI scores
- # of total mutations
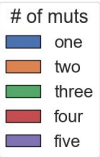- # of total oncogenic mutations

Mutations (1/2)

Number of tp53 mutations per sample

tp53_count	prim_count	met_count	sum
1	6549	5168	11717
2	593	304	897
3	67	29	96
4	14	3	17
5	3	1	4

- Strong enrichment for 1 TP53 mutation per sample

- Enrichment of 2 TP53 mutations in Primary Samples compared to Metastatic.
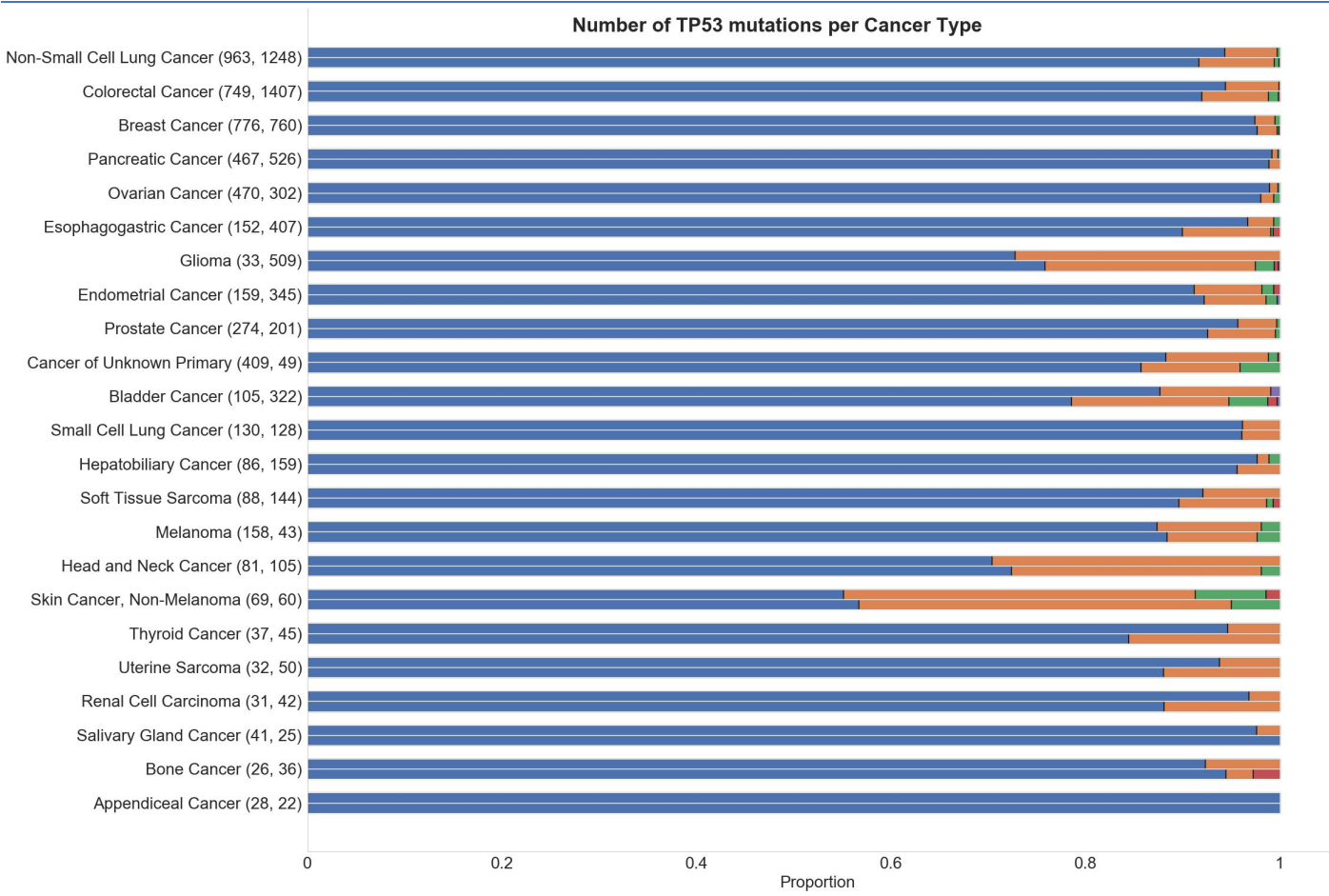
1st Line: Metastasis
2nd Line: Primary

TP53 State Study: FACETS Output

3 types of events at TP53 Locus:
- Whole Genome Duplication
- Copy Number Events
- Mutational Events

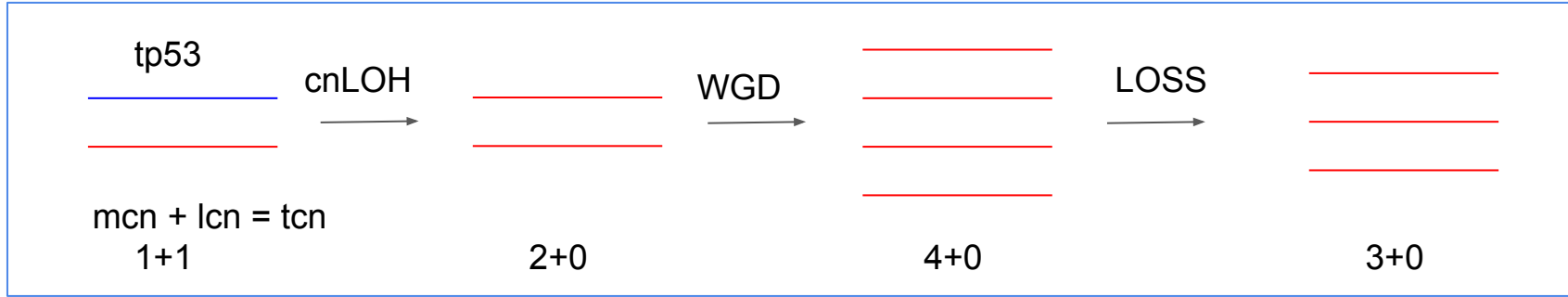Copy Number States from FACETS are chronological:

1. Copy Number Event → CNLOH / LOSS / GAIN / AMP / DIPLOID / TETRAPLOID

2. [Optional] WGD chronology → BEFORE / AFTER

3. [Optional] Copy Number Event → LOSS / GAIN / AMP
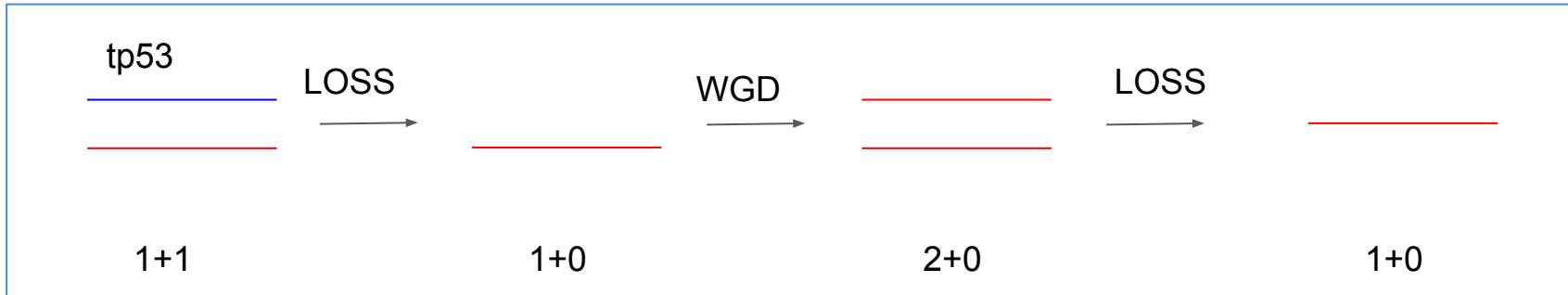
Examples:

DIPLOID   HETLOSS   CNLOH   HOMDEL
CNLOH BEFORE & LOSS   DOUBLE LOSS AFTER   LOSS BEFORE & AFTER

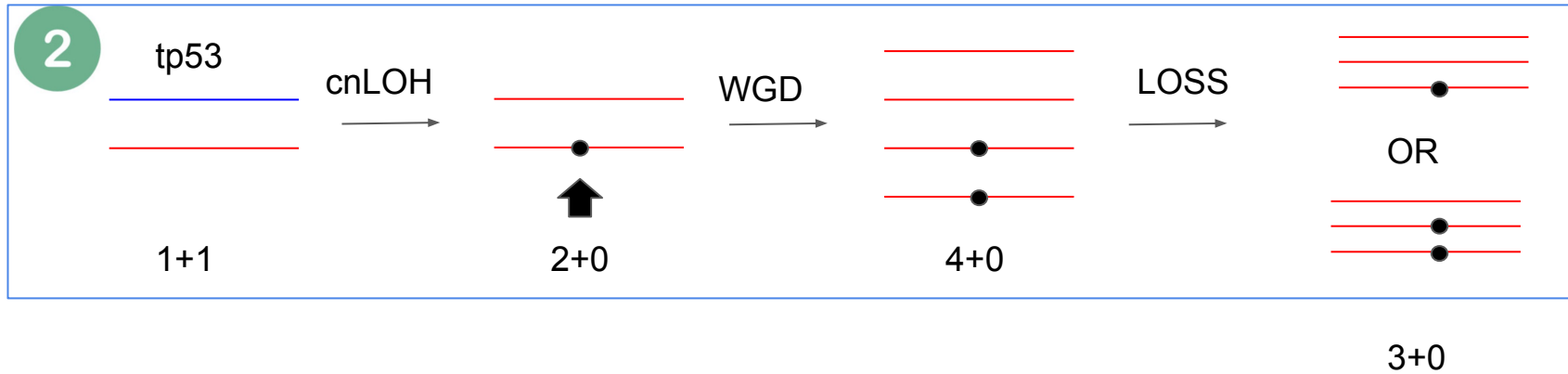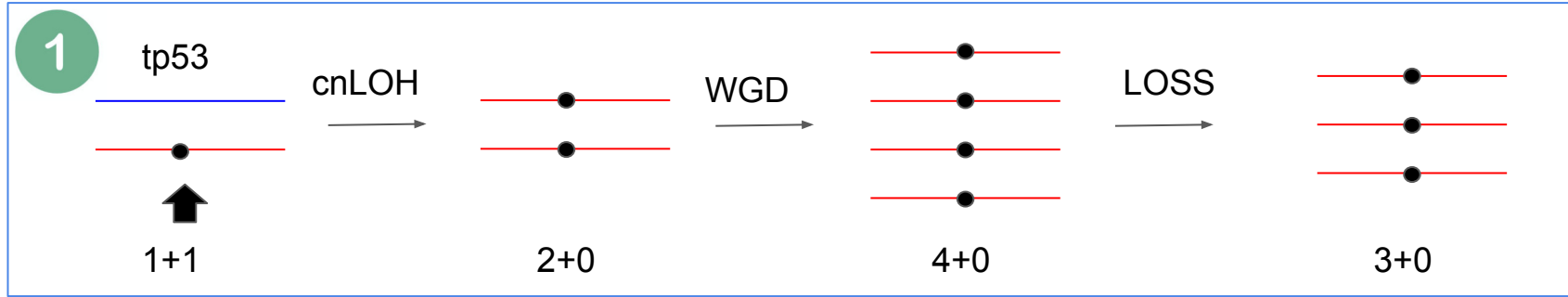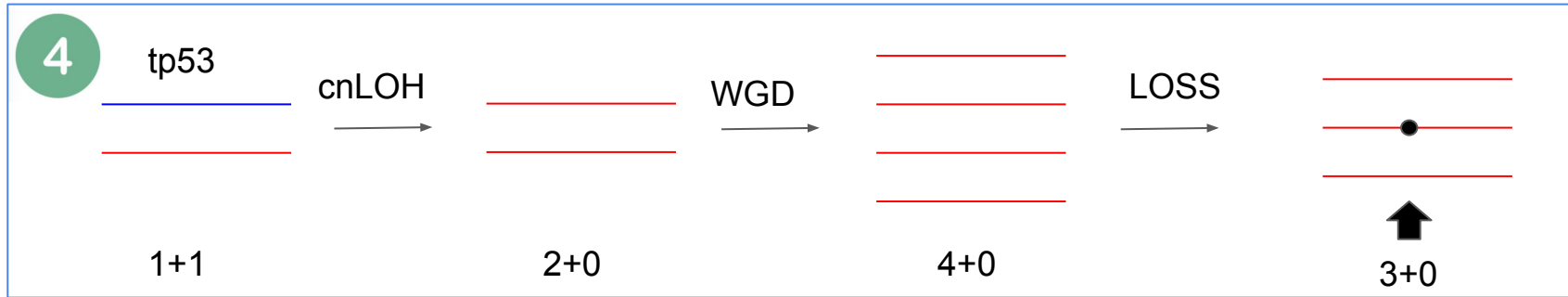# TP53 State Study: Allelic Representation

CNLOH BEFORE & LOSS

tp53

cnLOH → WGD → LOSS →

mcn + lcn = tcn

1+1   2+0   4+0   3+0

LOSS BEFORE & AFTER

tp53

LOSS → WGD → LOSS →

1+1   1+0   2+0   1+0
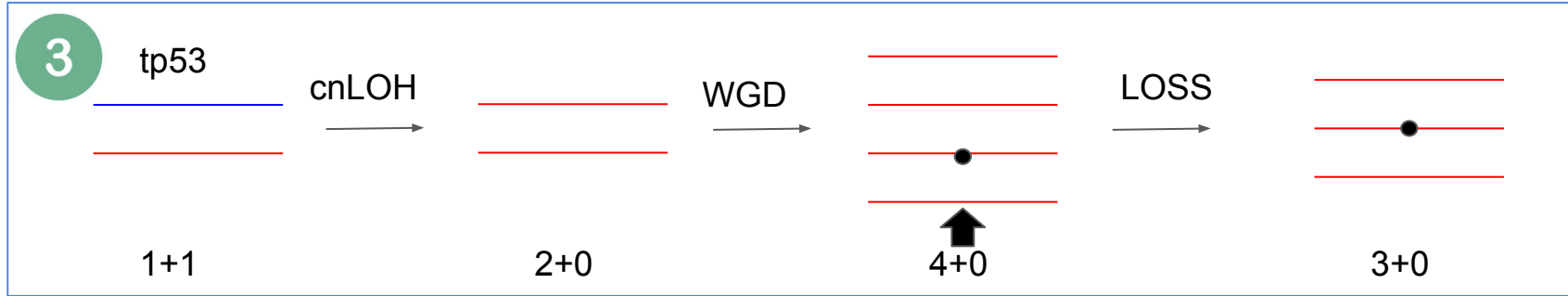
TP53 State Study: With mutation

TP53 State Study: With mutation

CNLOH BEFORE & LOSS

TP53 State Study: With mutation

Mutation Timing	# of tp53 mutant copies	# WT Alleles
mut before CNLOH	3	0
mut after CNLOH, before WGD	1	2
mut after CNLOH, before WGD	2	1
mut after WGD, before LOSS	1	2
mut after LOSS	1	2

TP53 State Study: With mutation

Mutation Timing	# of tp53 mutant copies	# WT Alleles
mut before CNLOH	3	0
mut after CNLOH, before WGD	1	2
mut after CNLOH, before WGD	2	1
mut after WGD, before LOSS	1	2
mut after LOSS	1	2

Objective: Define groups based on similar Copy Number Events and on similar TP53 allelic state

Idea: The number of tp53 mutant copies allows us to know the number of WT alleles because **# of WT = TCN - # of mutants**.
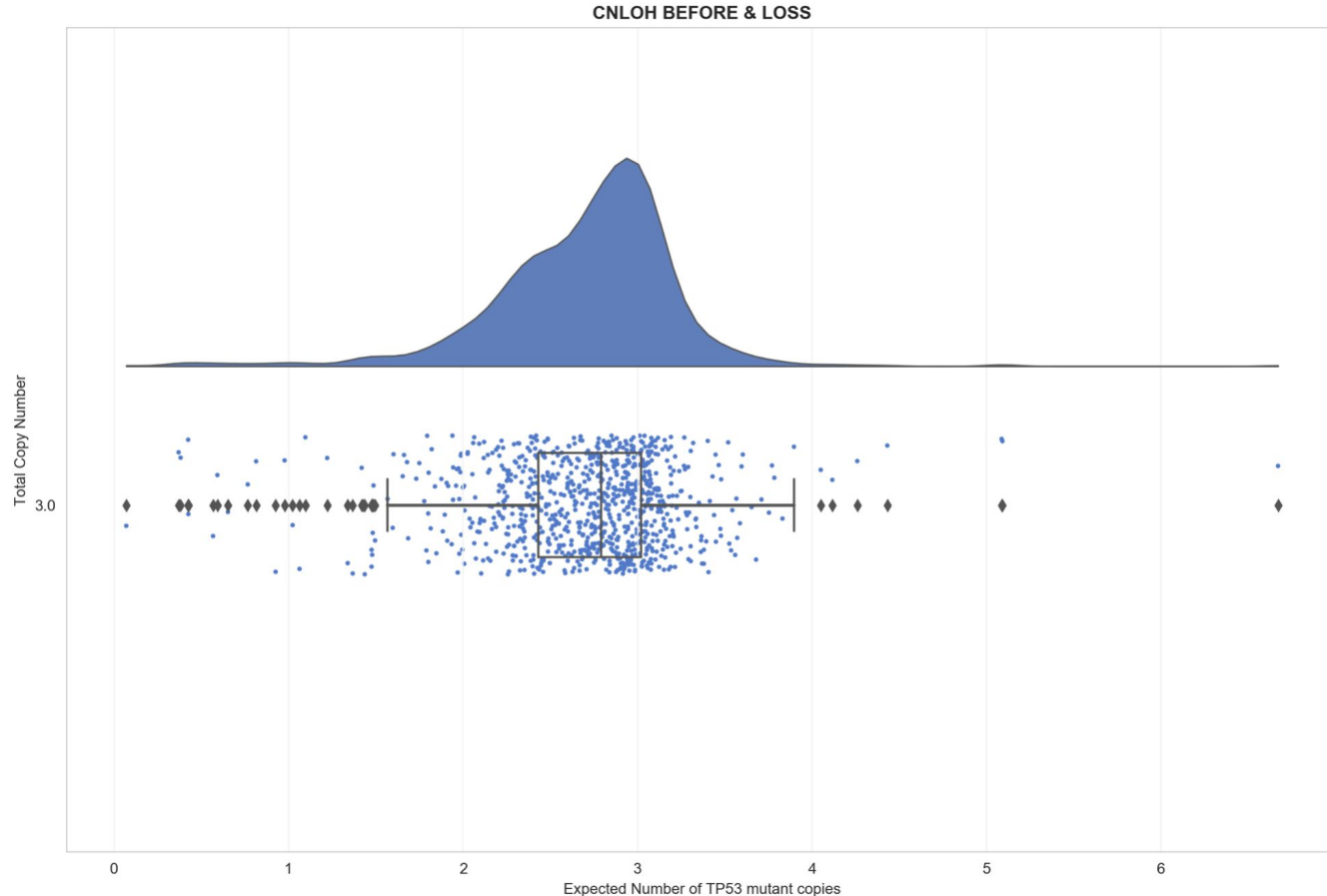
→ **Plot the expected number of tp53 mutant copies against the TCN**

$$\frac{vaf}{purity} \times (tcn \times purity + 2 \times (1 - purity)))$$

https://www.nature.com/articles/s41588-018-0165-1

TP53 State Study: Interpretation

CNLOH BEFORE & LOSS

- The peak of the distribution is reached for exp_nb ≈ 3.

- So # WT = tcn - exp_nb
  $$= 3 - 3$$
  $$= 0$$

# Definition of TP53 subgroups

First Look at the distributions

We grouped the different Copy Number States into 4 groups to look at the data and compare the distributions:

- **<u>CNLOH</u>** = ['CNLOH', 'CNLOH BEFORE & LOSS', 'CNLOH AFTER', 'CNLOH BEFORE', 'CNLOH & GAIN', 'CNLOH BEFORE & GAIN', 'AMP (LOH)']

- **<u>LOSS</u>** = ['HETLOSS', 'LOSS BEFORE', 'LOSS AFTER', 'HOMDEL', 'LOSS BEFORE & AFTER', 'DOUBLE LOSS AFTER']

- **<u>GAIN</u>** = ['GAIN', 'AMP', 'AMP (BALANCED)', 'LOSS & GAIN']

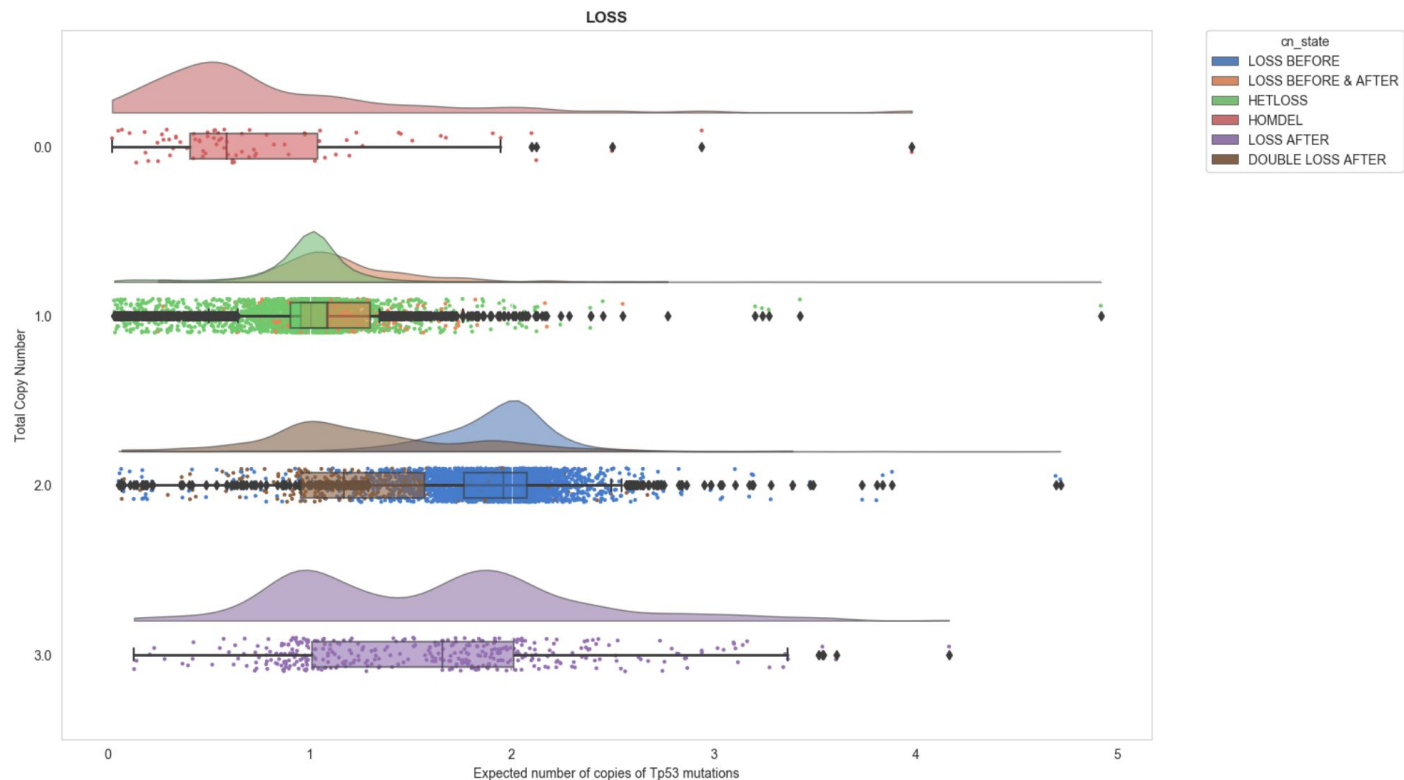- **<u>DIPLOID & TETRAPLOID</u>** = ['DIPLOID', 'TETRAPLOID']

CNLOH

cn_state	count
AMP (LOH)	12
CNLOH	802
CNLOH & GAIN	61
CNLOH AFTER	43
CNLOH BEFORE	174
CNLOH BEFORE & GAIN	26
CNLOH BEFORE & LOSS	893
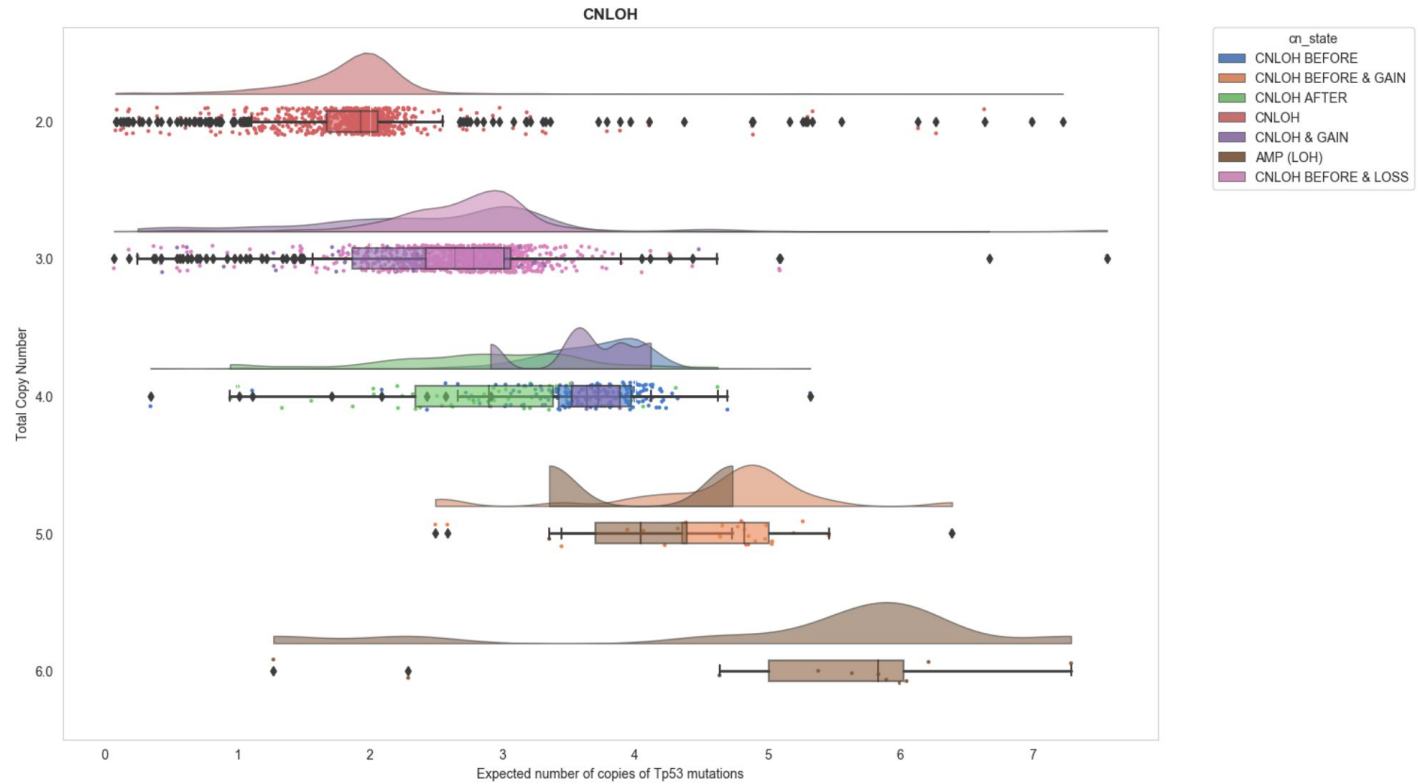
LOSS

cn_state	count
DOUBLE LOSS AFTER	149
HETLOSS	3727
HOMDEL	71
LOSS AFTER	271
LOSS BEFORE	3131
LOSS BEFORE & AFTER	226

DIPLOID & TETRAPLOID
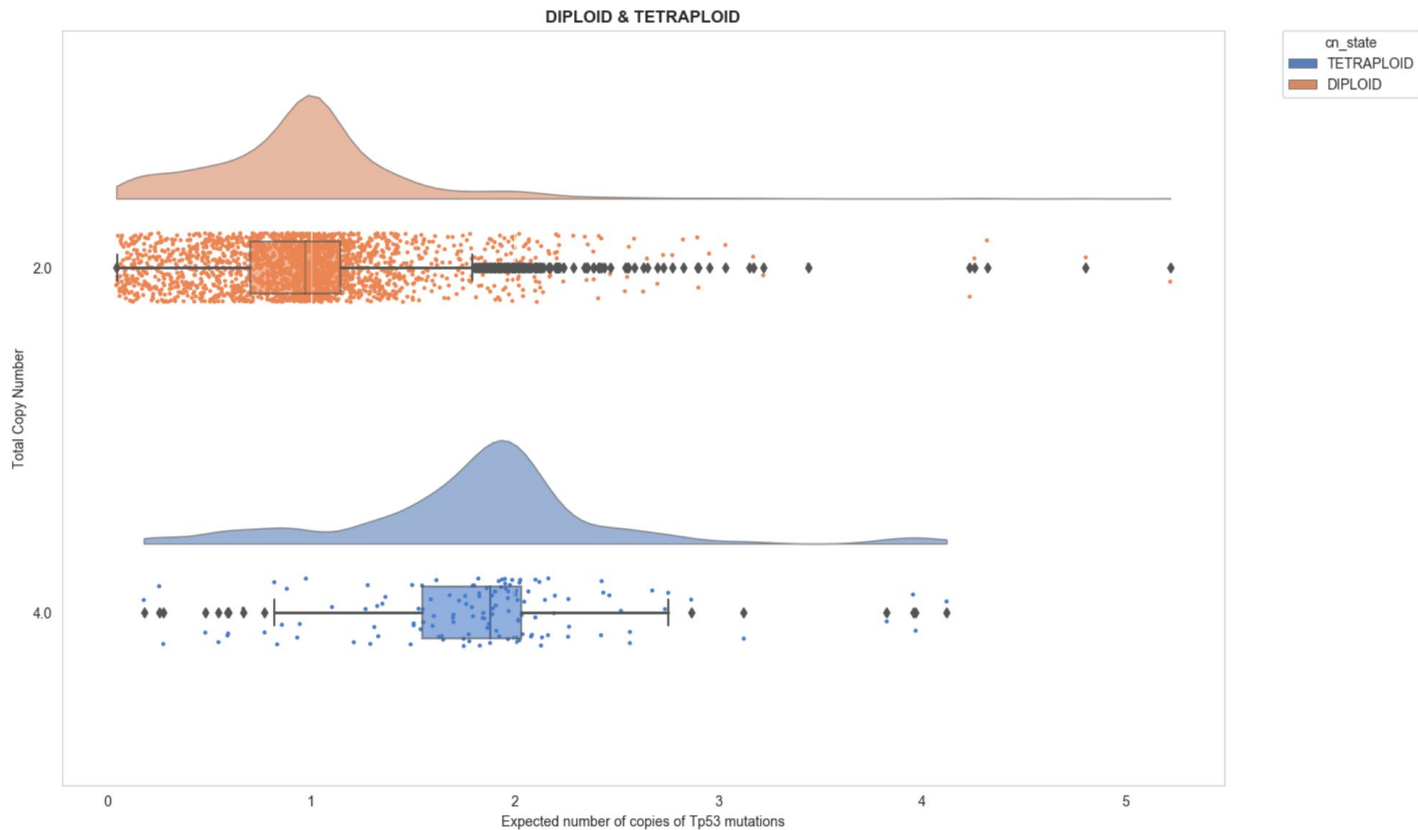
GAIN

cn_state	count
AMP	12
AMP (BALANCED)	4
GAIN	91
LOSS & GAIN	15
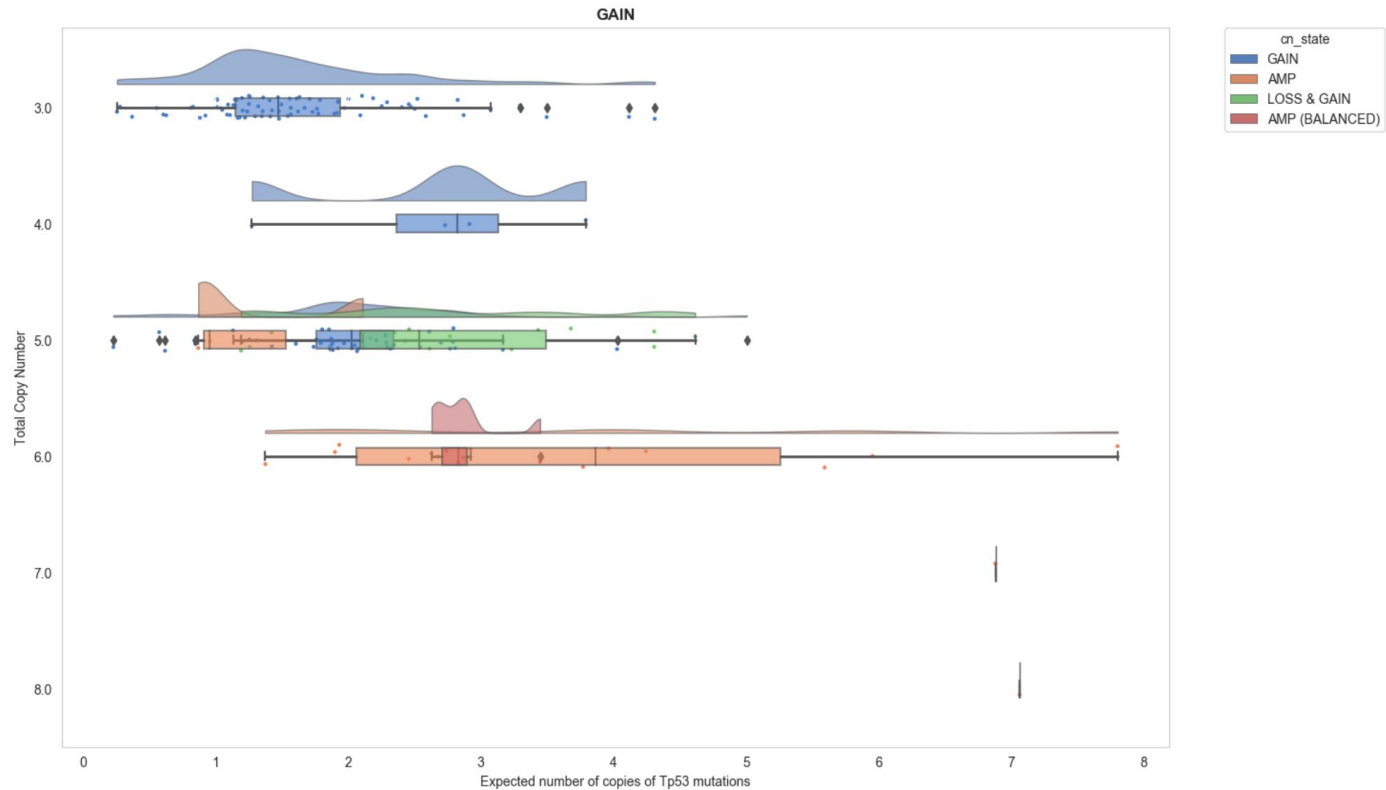
Subgroups Definition

No Allele groups:
- HOMDEL
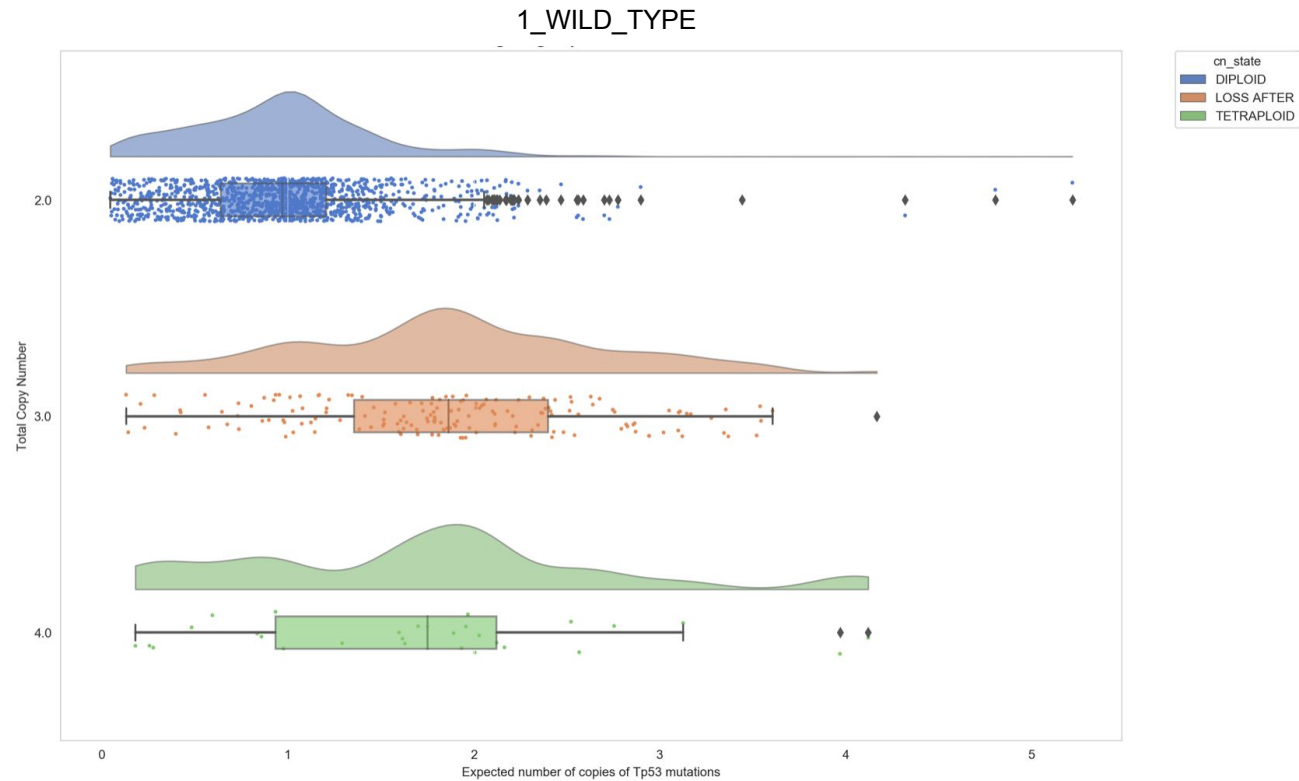
Mono-allelic groups (WT allele is still present):
- Reference group: 0 mutation and a LOSS
- TP53 mutant group: 1 TP53 mutation and DIPLOID / TETRAPLOID / LOSS AFTER

Possibly bi-allelic groups:
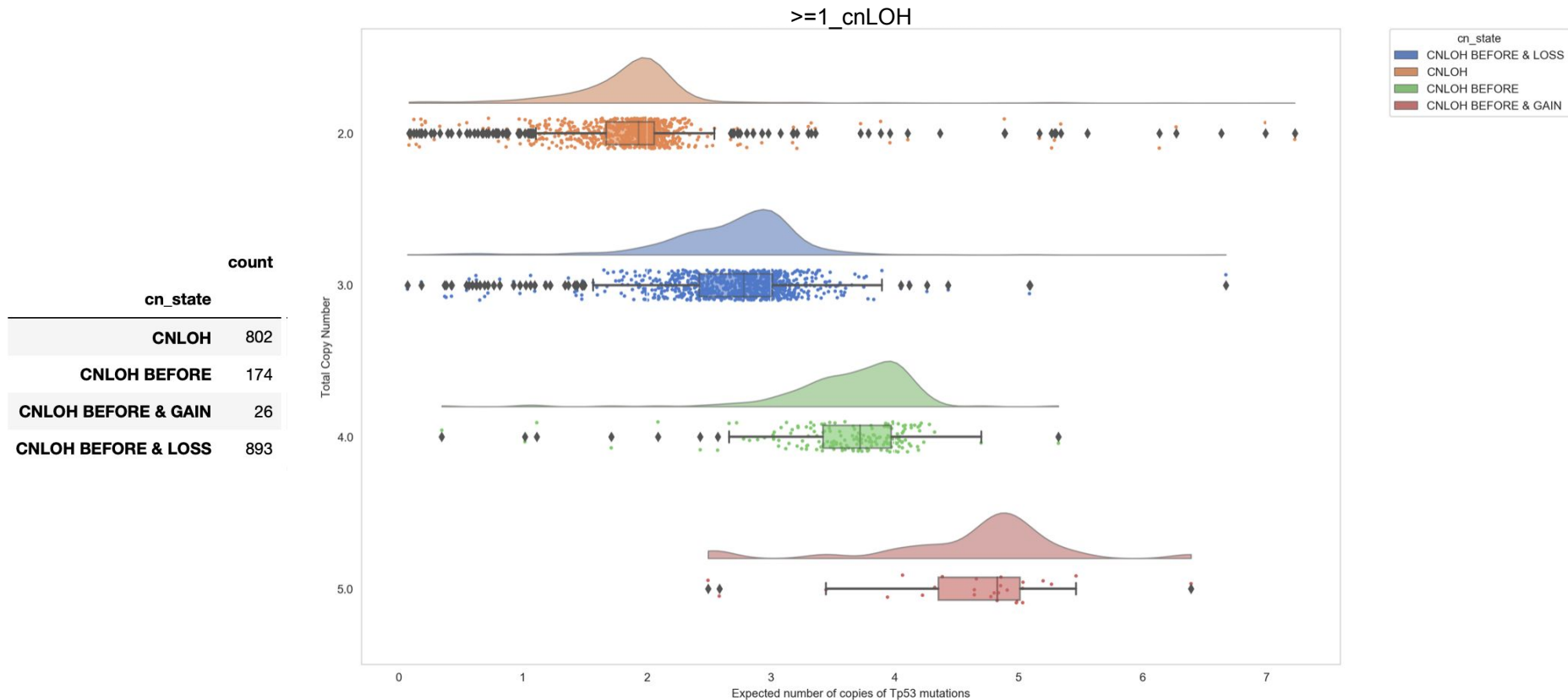- cnLOH group: at least 1 TP53 mutation and a CNLOH event
- Loss group: at least 1 TP53 mutation and a LOSS event LOSS BEFORE / HETLOSS / LOSS BEFORE & AFTER
- >1 mut group : at least 2 TP53 mutations and not in Loss or cnLOH groups

1_WILD_TYPE

cn_state	count
DIPLOID	1262
LOSS AFTER	166
TETRAPLOID	29

>=1_cnLOH

>=1_LOSS

>=1_LOSS

>1muts

mut_cn_group	count
>=1_LOSS	7087
0_HETLOSS	2833
>=1_cnLOH	1895
1_WILD_TYPE	1457
>1muts	697
HOMDEL	296
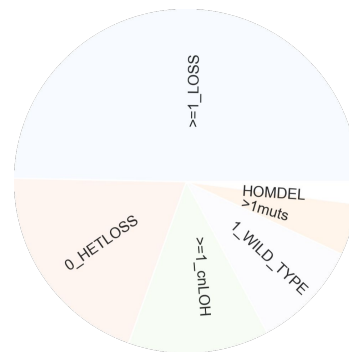
mut_cn_group	tp53_count	count
0_HETLOSS	0	2833
1_WILD_TYPE	1	1457
>1muts	2	616
	3	68
	4	10
	5	3
>=1_LOSS	1	6904
	2	161
	3	19
	4	3
>=1_cnLOH	1	1850
	2	39
	3	6
HOMDEL	0	224
	1	63
	2	8
	4	1
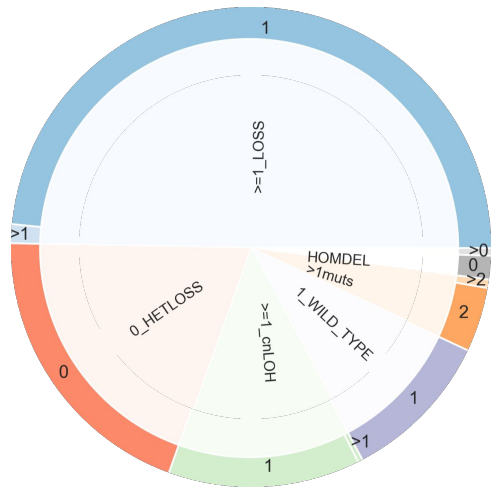
mut_cn_group	Sample_Type	count
0_HETLOSS	Metastasis	1260
	Primary	1556
1_WILD_TYPE	Metastasis	494
	Primary	963
>1muts	Metastasis	232
	Primary	465
>=1_LOSS	Metastasis	3271
	Primary	3816
>=1_cnLOH	Metastasis	821
	Primary	1074
HOMDEL	Metastasis	116
	Primary	179
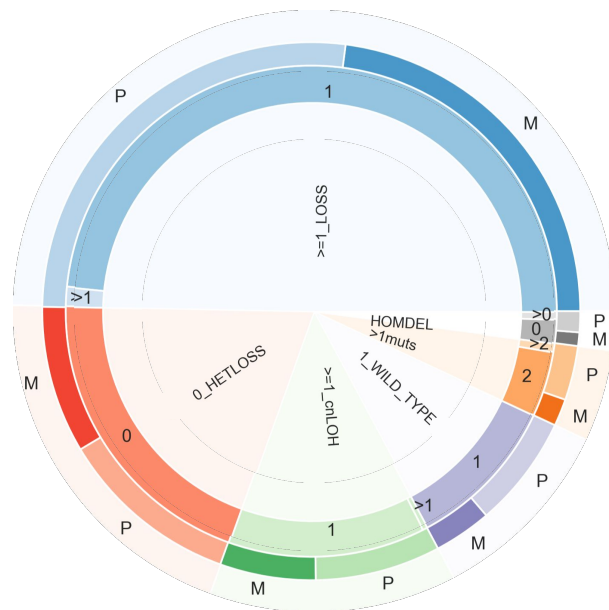
mut_cn_group	Cancer_Type	count
0_HETLOSS	**Breast Cancer**	864
	Non-Small Cell Lung Cancer	347
	Prostate Cancer	254
1_WILD_TYPE	**Non-Small Cell Lung Cancer**	322
	Colorectal Cancer	198
	Pancreatic Cancer	142
>1muts	**Colorectal Cancer**	116
	Non-Small Cell Lung Cancer	110
	Glioma	101
>=1_LOSS	**Colorectal Cancer**	1361
	Non-Small Cell Lung Cancer	1158
	Breast Cancer	1005
>=1_cnLOH	**Non-Small Cell Lung Cancer**	348
	Colorectal Cancer	281
	Breast Cancer	237
HOMDEL	**Soft Tissue Sarcoma**	47
	Breast Cancer	41
	Prostate Cancer	34

Subgroups Summary (2/2)

Distribution of subgroups for each Cancer Type

Enrichments

- 0_HETLOSS :
  - Breast Cancer,
  - Mesothelomia

- \>1muts:
  - Skin Cancer,
  - Head and Neck Cancer,
  - Glioma
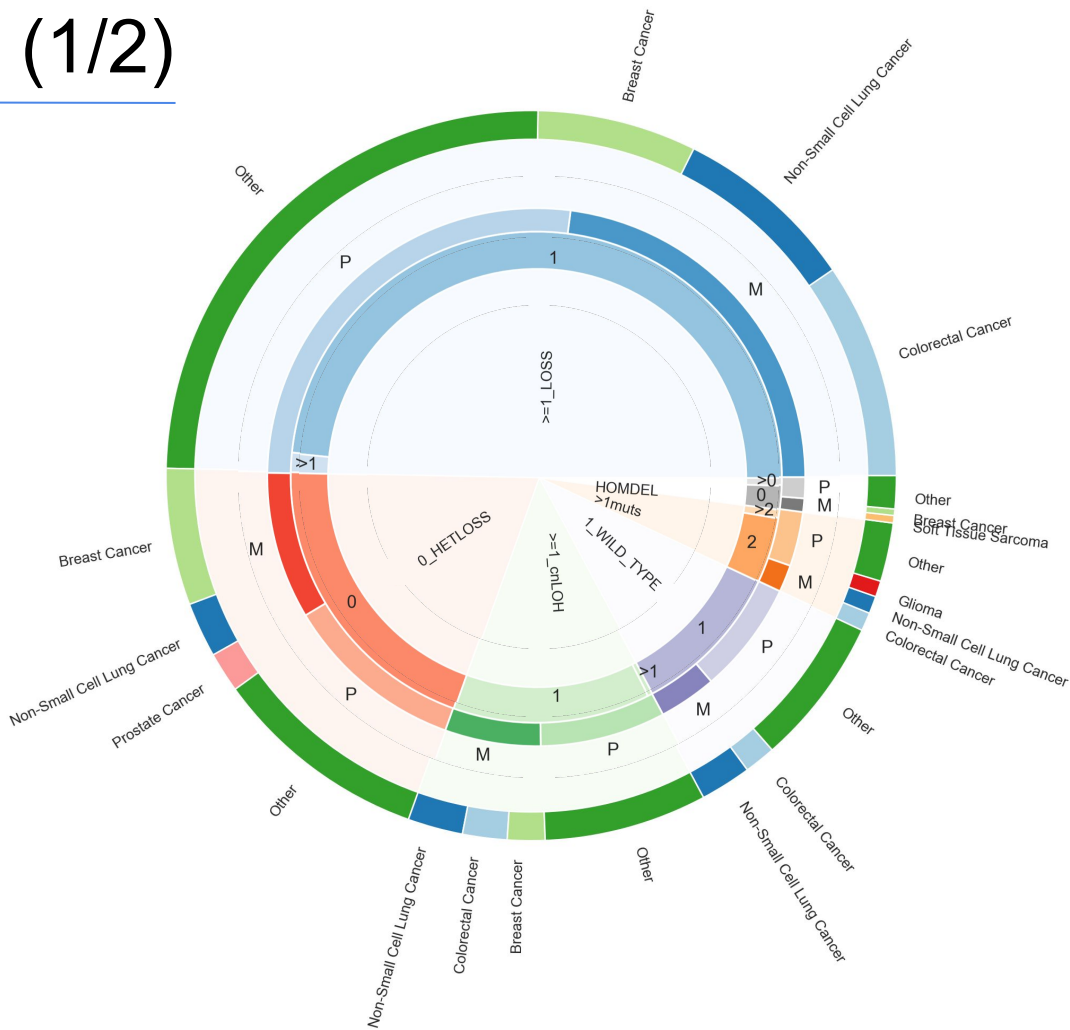
- \>=1_cnLOH:
  - Glioma
  - Germ Cell Tumor
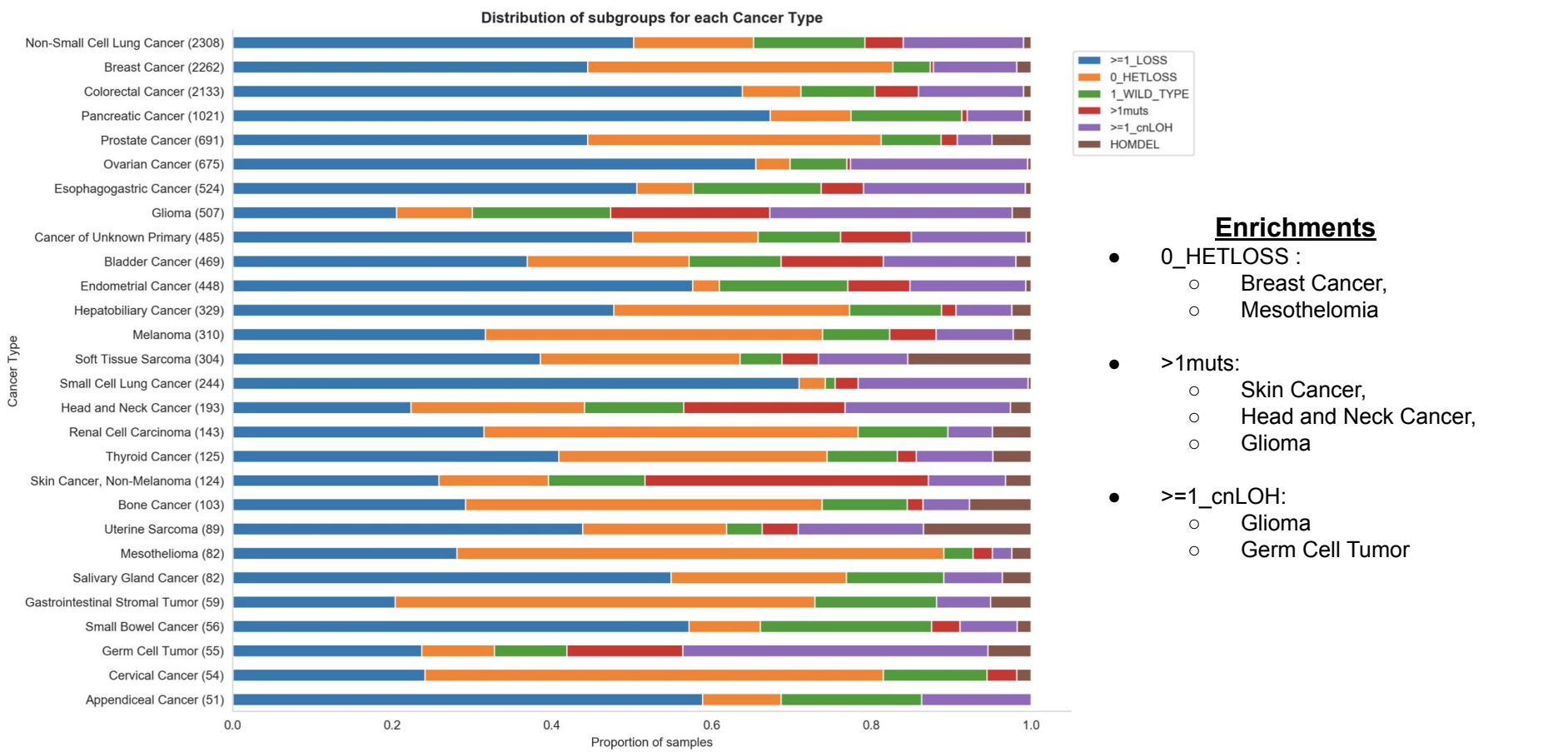
Correlative Analysis

Data Cleaning: Purity

Data Cleaning: Purity

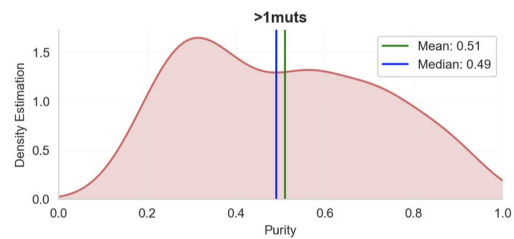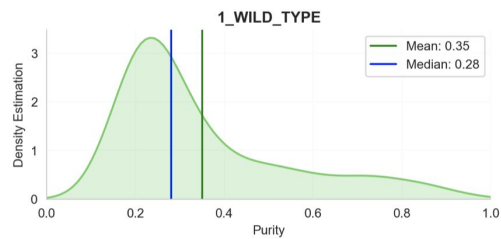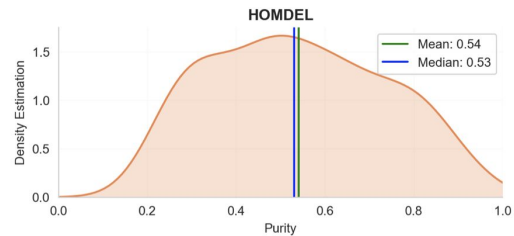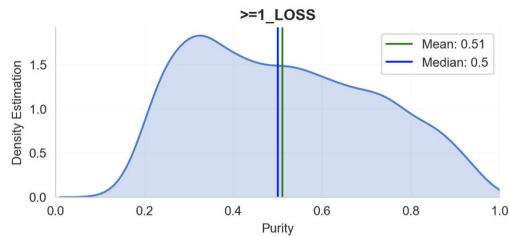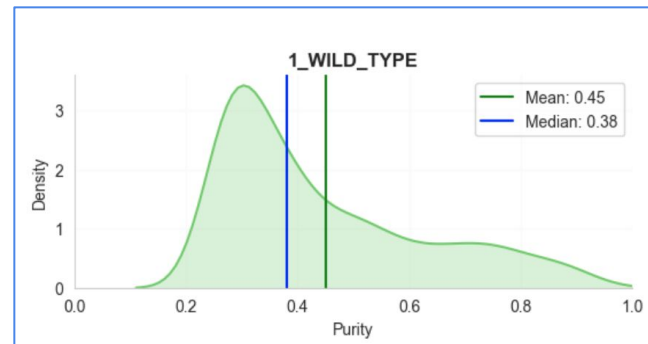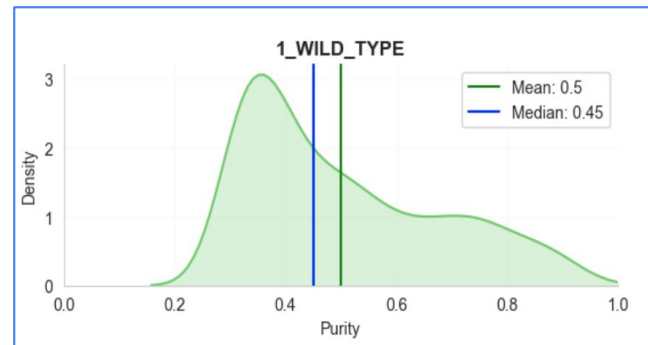|	>=1_cnLOH	1_WILD_TYPE	>1muts	HOMDEL	>=1_LOSS	0_HETLOSS	SUM	ratio	purity_median_interval
0.20	70	239	39	4	135	57	544	3.8	[0.32, 0.52]
0.21	98	303	52	4	210	85	752	5.3	[0.33, 0.52]
0.22	117	377	62	6	323	116	1001	7.0	[0.34, 0.52]
0.23	135	443	74	13	429	149	1243	8.7	[0.36, 0.53]
0.24	174	497	86	16	535	191	1499	10.5	[0.37, 0.54]
0.25	219	553	95	19	657	237	1780	12.5	[0.38, 0.54]
0.26	245	600	106	21	807	287	2066	14.5	[0.39, 0.54]
0.27	282	655	131	27	937	344	2376	16.7	[0.41, 0.55]
0.28	323	709	151	38	1062	383	2666	18.7	[0.42, 0.55]
0.30	405	785	183	43	1355	470	3241	22.7	[0.45, 0.56]
0.31	437	824	194	45	1504	520	3524	24.7	[0.47, 0.57]
0.32	478	856	207	48	1679	577	3845	27.0	[0.49, 0.57]
0.33	512	895	216	54	1851	616	4144	29.1	[0.5, 0.58]
0.34	550	916	232	58	1995	663	4414	30.9	[0.51, 0.58]
0.35	593	936	246	60	2159	719	4713	33.0	[0.52, 0.59]
0.36	632	958	264	66	2302	765	4987	35.0	[0.53, 0.59]
0.37	676	985	282	72	2441	808	5264	36.9	[0.54, 0.6]
0.40	773	1046	309	92	2841	941	6002	42.1	[0.58, 0.62]
0.60	1372	1268	486	192	5059	1831	10208	71.6	[0.73, 0.76]
0.80	1739	1403	616	267	6480	2526	13031	91.3	[0.85, 0.87]
1.00	1895	1457	697	296	7087	2833	14265	100.0	[nan, nan]

- Cut: 0.25
- Population lost: 12.5 %
- Medians: [0.38, 0.54]
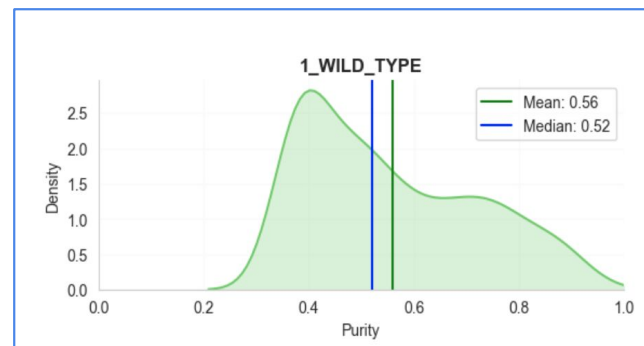
Data Cleaning: Purity

	>=1_cnLOH	1_WILD_TYPE	>1muts	HOMDEL	>=1_LOSS	0_HETLOSS	SUM	ratio	purity_median_interval
0.20	70	239	39	4	135	57	544	3.8	[0.32, 0.52]
0.21	98	303	52	4	210	85	752	5.3	[0.33, 0.52]
0.22	117	377	62	6	323	116	1001	7.0	[0.34, 0.52]
0.23	135	443	74	13	429	149	1243	8.7	[0.36, 0.53]
0.24	174	497	86	16	535	191	1499	10.5	[0.37, 0.54]
0.25	219	553	95	19	657	237	1780	12.5	[0.38, 0.54]
0.26	245	600	106	21	807	287	2066	14.5	[0.39, 0.54]
0.27	282	655	131	27	937	344	2376	16.7	[0.41, 0.55]
0.28	323	709	151	38	1062	383	2666	18.7	[0.42, 0.55]
0.30	405	785	183	43	1355	470	3241	22.7	[0.45, 0.56]
0.31	437	824	194	45	1504	520	3524	24.7	[0.47, 0.57]
0.32	478	856	207	48	1679	577	3845	27.0	[0.49, 0.57]
0.33	512	895	216	54	1851	616	4144	29.1	[0.5, 0.58]
0.34	550	916	232	58	1995	663	4414	30.9	[0.51, 0.58]
0.35	593	936	246	60	2159	719	4713	33.0	[0.52, 0.59]
0.36	632	958	264	66	2302	765	4987	35.0	[0.53, 0.59]
0.37	676	985	282	72	2441	808	5264	36.9	[0.54, 0.6]
0.40	773	1046	309	92	2841	941	6002	42.1	[0.58, 0.62]
0.60	1372	1268	486	192	5059	1831	10208	71.6	[0.73, 0.76]
0.80	1739	1403	616	267	6480	2526	13031	91.3	[0.85, 0.87]
1.00	1895	1457	697	296	7087	2833	14265	100.0	[nan, nan]

- Cut: 0.30
- Population lost: 22.7 %
- Medians: [0.45, 0.56]

Data Cleaning: Purity

	>=1_cnLOH	1_WILD_TYPE	>1muts	HOMDEL	>=1_LOSS	0_HETLOSS	SUM	ratio	purity_median_interval
0.20	70	239	39	4	135	57	544	3.8	[0.32, 0.52]
0.21	98	303	52	4	210	85	752	5.3	[0.33, 0.52]
0.22	117	377	62	6	323	116	1001	7.0	[0.34, 0.52]
0.23	135	443	74	13	429	149	1243	8.7	[0.36, 0.53]
0.24	174	497	86	16	535	191	1499	10.5	[0.37, 0.54]
0.25	219	553	95	19	657	237	1780	12.5	[0.38, 0.54]
0.26	245	600	106	21	807	287	2066	14.5	[0.39, 0.54]
0.27	282	655	131	27	937	344	2376	16.7	[0.41, 0.55]
0.28	323	709	151	38	1062	383	2666	18.7	[0.42, 0.55]
0.30	405	785	183	43	1355	470	3241	22.7	[0.45, 0.56]
0.31	437	824	194	45	1504	520	3524	24.7	[0.47, 0.57]
0.32	478	856	207	48	1679	577	3845	27.0	[0.49, 0.57]
0.33	512	895	216	54	1851	616	4144	29.1	[0.5, 0.58]
0.34	550	916	232	58	1995	663	4414	30.9	[0.51, 0.58]
0.35	593	936	246	60	2159	719	4713	33.0	[0.52, 0.59]
0.36	632	958	264	66	2302	765	4987	35.0	[0.53, 0.59]
0.37	676	985	282	72	2441	808	5264	36.9	[0.54, 0.6]
0.40	773	1046	309	92	2841	941	6002	42.1	[0.58, 0.62]
0.60	1372	1268	486	192	5059	1831	10208	71.6	[0.73, 0.76]
0.80	1739	1403	616	267	6480	2526	13031	91.3	[0.85, 0.87]
1.00	1895	1457	697	296	7087	2833	14265	100.0	[nan, nan]

- Cut: 0.35
- Population lost: 33.0 %
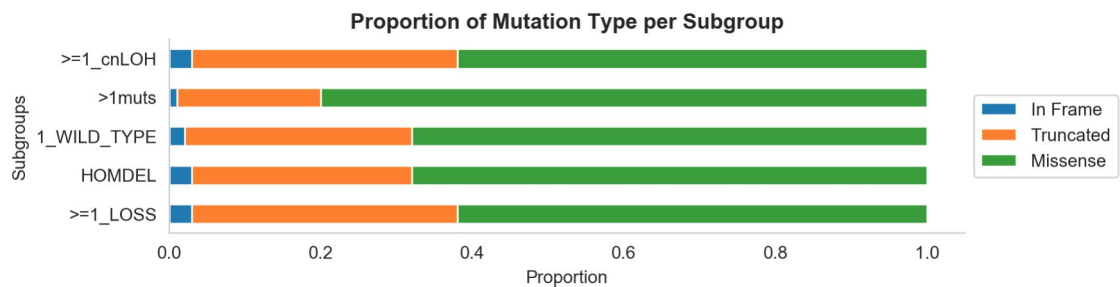- Medians: [0.52, 0.59]

Study Plan

- >1muts Subgroup Exploration

- Genome-wide genomic instability (# of chr loh, loss, gain)

- Tumor Burden and co-mutations (TMB, MSI, Oncogenic)

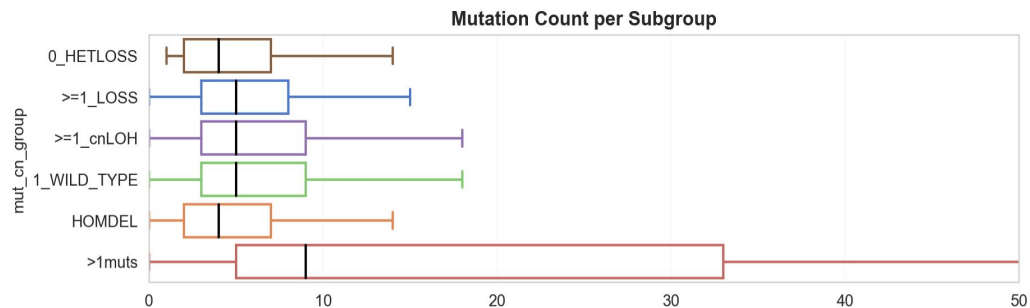- Clinical correlates when available (survival …)

>1muts Subgroup
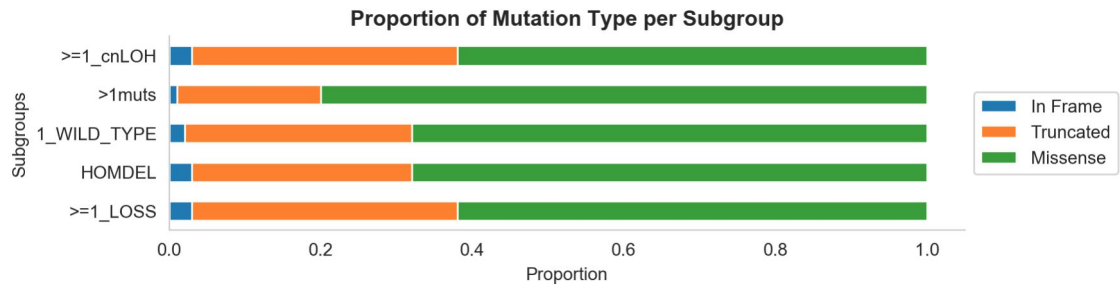
Proportion of Mutation Type per Subgroup

- Enrichment of Missense Mutations in >1muts Subgroup

>1muts Subgroup

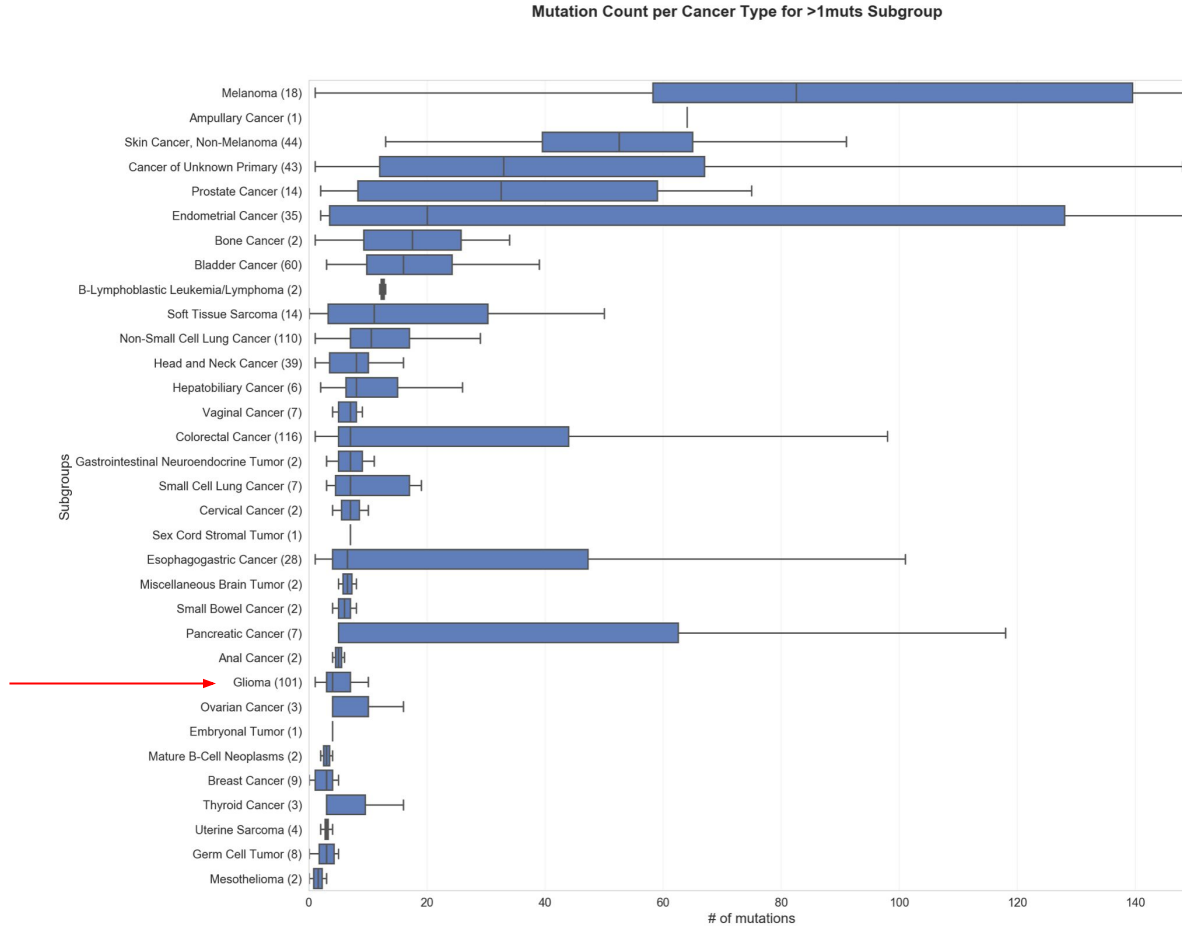Proportion of Mutation Type per Subgroup

Mutation Count per Subgroup

- ● Enrichment of Missense Mutations in >1muts Subgroup

- ● Higher mutation count across all IMPACT genes in >1muts subgroup

>1muts Subgroup

Mutation Count per Cancer Type for >1muts Subgroup

Thank You