

Hematopoietic and Lymphoid System Neoplasms

Survival Analysis

BIO392 - Bioinformatics of Sequence Variation

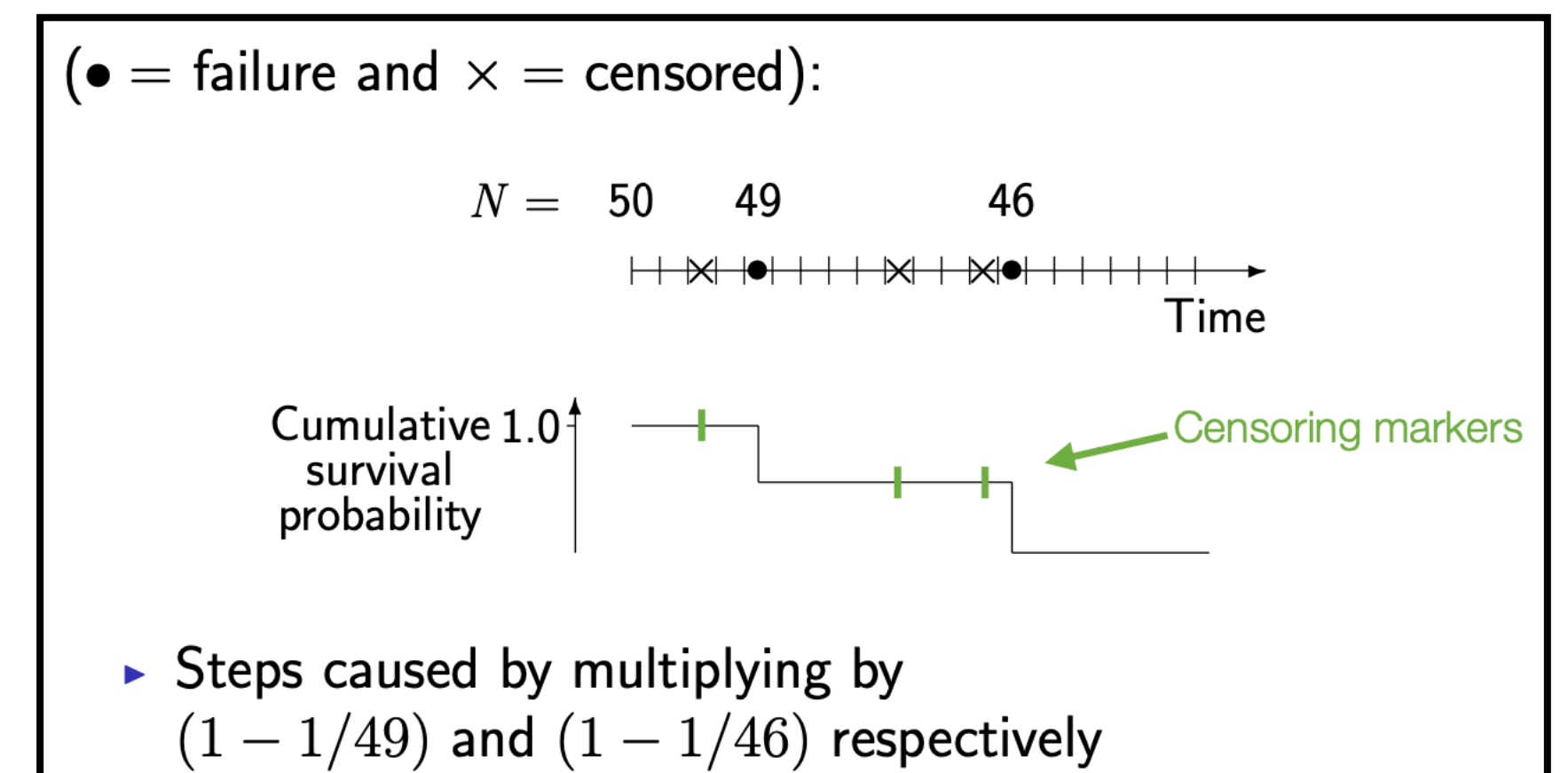
Background

Background: Survival

Kaplan-Meier Analysis of Survival Based on Conditional Probabilities

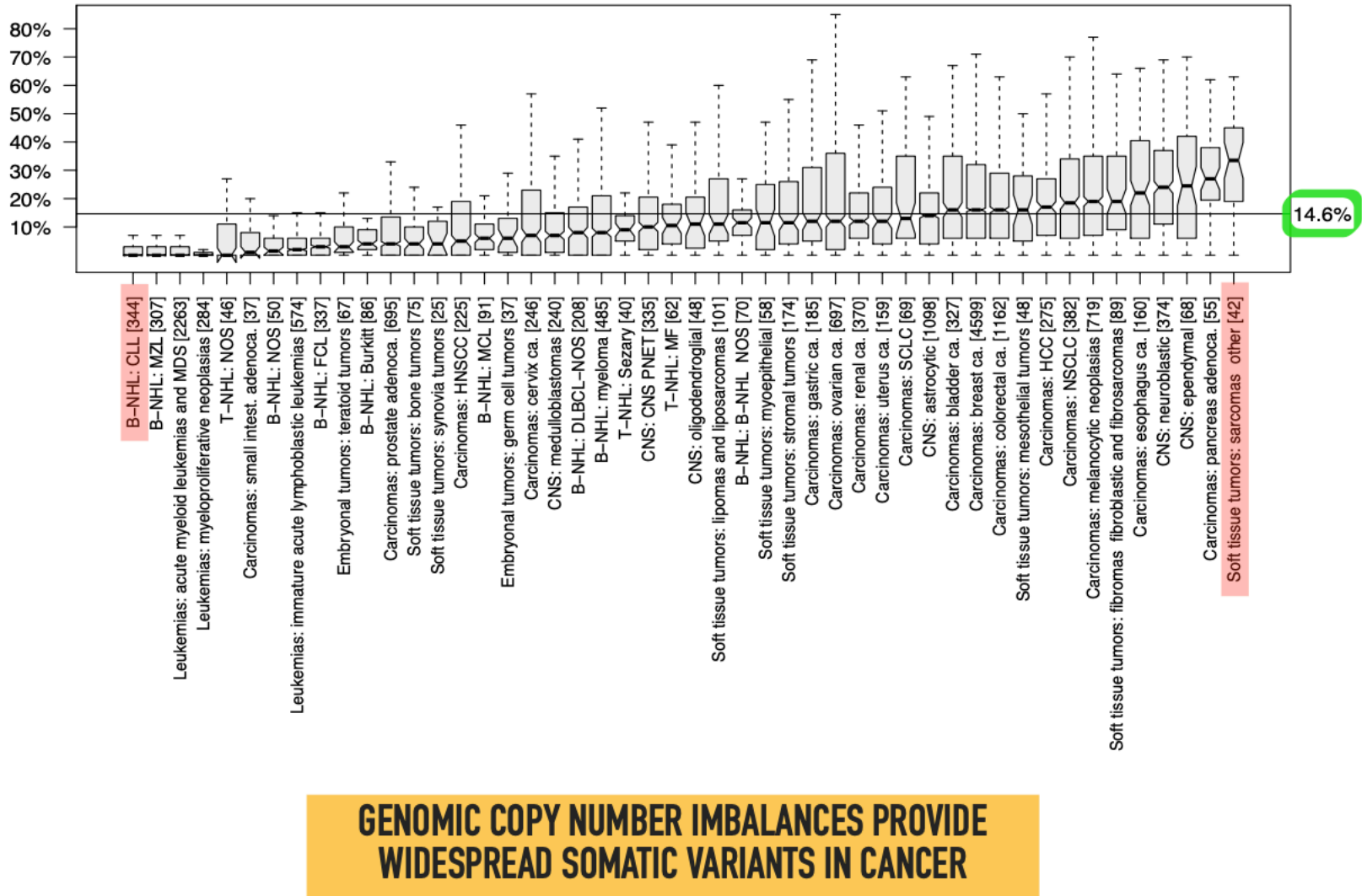
The Kaplan-Meier method

- Most common method to estimate survival function
- Based on conditional probabilities as we are interested in the probability of a subject surviving the next time interval given that they have survived so far
- Censoring markers = death events



Background: Genomic Data

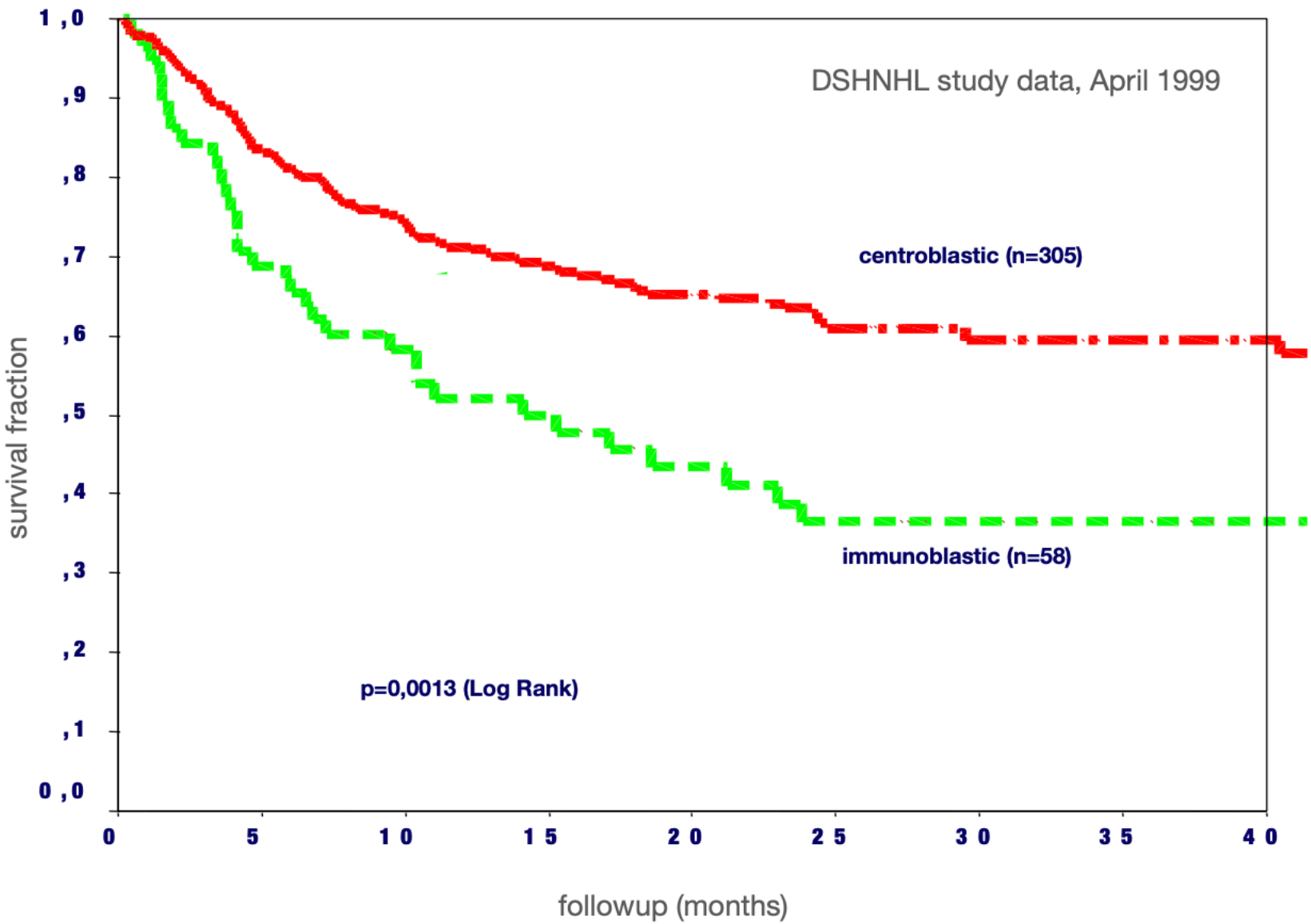
- **Copy number variation (CNV):**
nr of copies of a gene varies with respect to reference genome
- CNVs are common mutations in **cancer**



On average ~15% of a cancer genome are in an imbalanced state (more/less than 2 alleles);
Original data based on >30'000 cancer genomes from arraymap.org

- Cancer CNV frequencies correspond to diagnostic subsets: Kaplan-Meier Plots to visualize different risks

Difference in survival rate between different cancer patient subsets



Objectives

1) Predict survival in patients having cancer of the blood or bone marrow (e.g. leukemia, lymphoma, etc.)
using CNV sample data

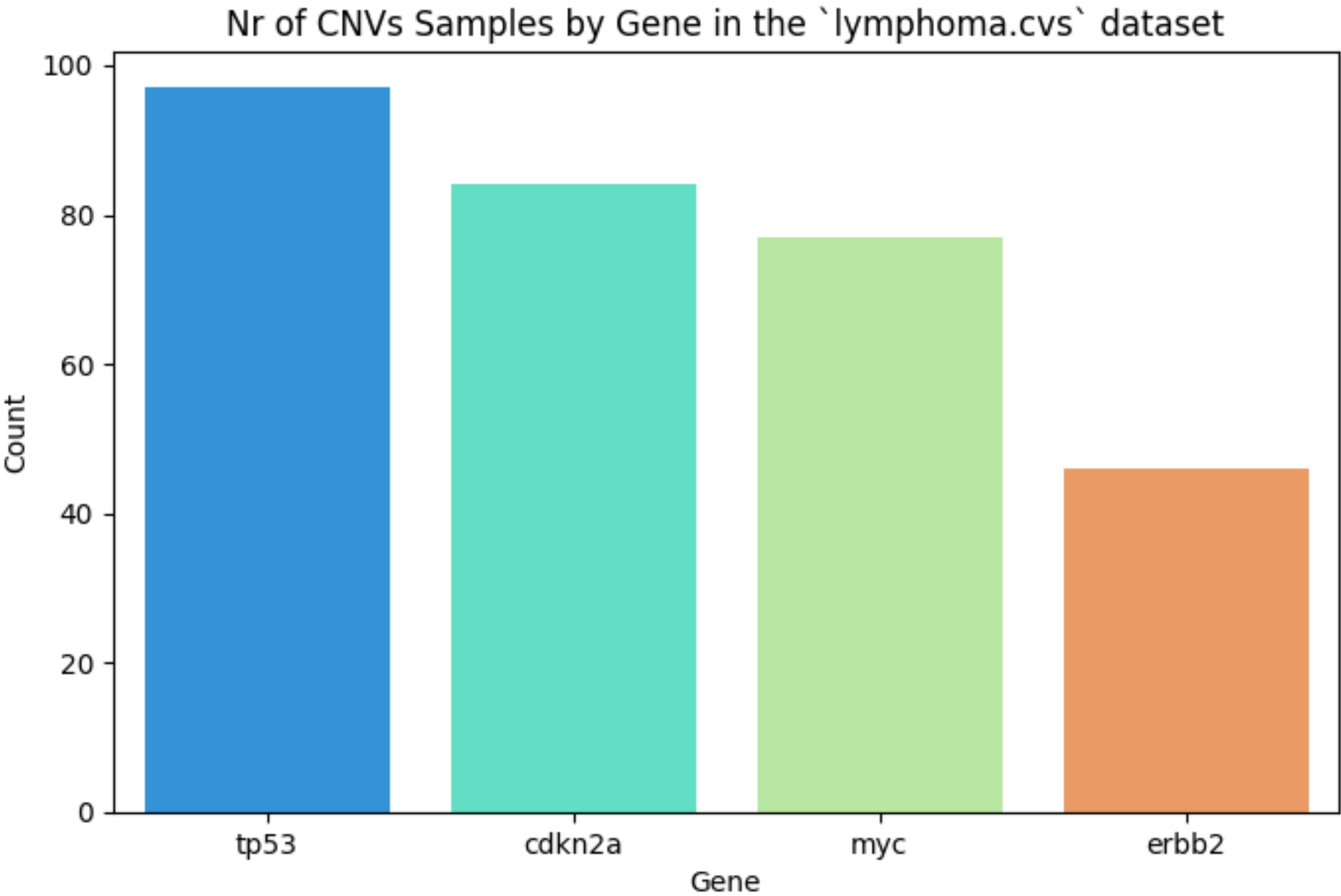
2) Identify variables that affect CNV fraction in the tumor samples

Additionally:

account for mutations in known tumor suppressor genes (TP53-, CDKN2A-)
and oncogenes (ERBB2+, MYC+)

Dataset

Cancer genome data: lymphoma.csv



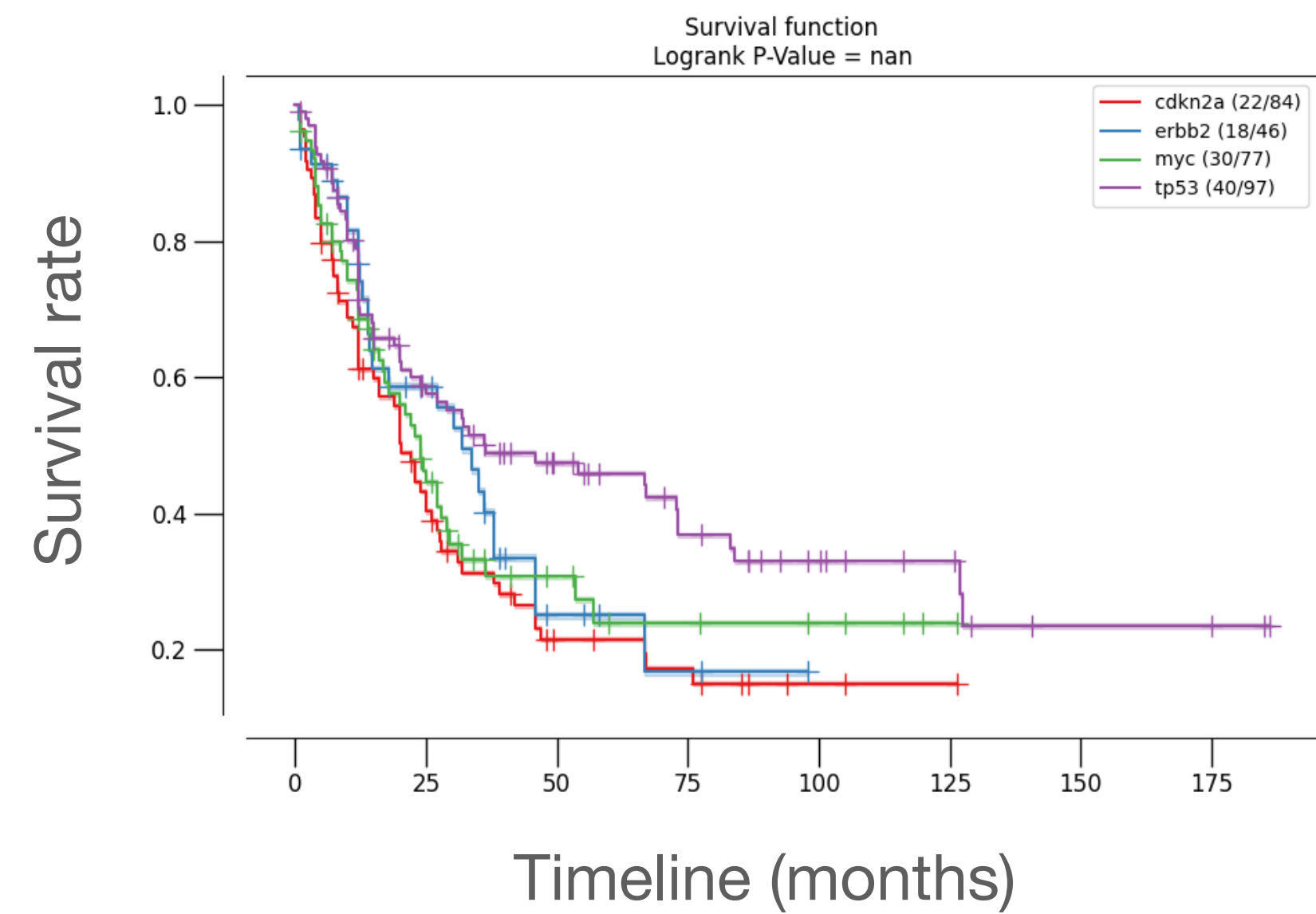
After matching...

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 237 entries, 4 to 48
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   info.followupMonths                  237 non-null   float64
1   info.death                          237 non-null   int64
2   group                               237 non-null   object
3   histologicalDiagnosis.id            237 non-null   object
4   info.cnvstatistics.cnvfraction      237 non-null   float64
5   sex                                 237 non-null   object
6   pathologicalStage.label             237 non-null   object
7   info.cnvstatistics.dupfraction      237 non-null   float64
8   info.cnvstatistics.delfraction      237 non-null   float64
dtypes: float64(4), int64(1), object(4)
```

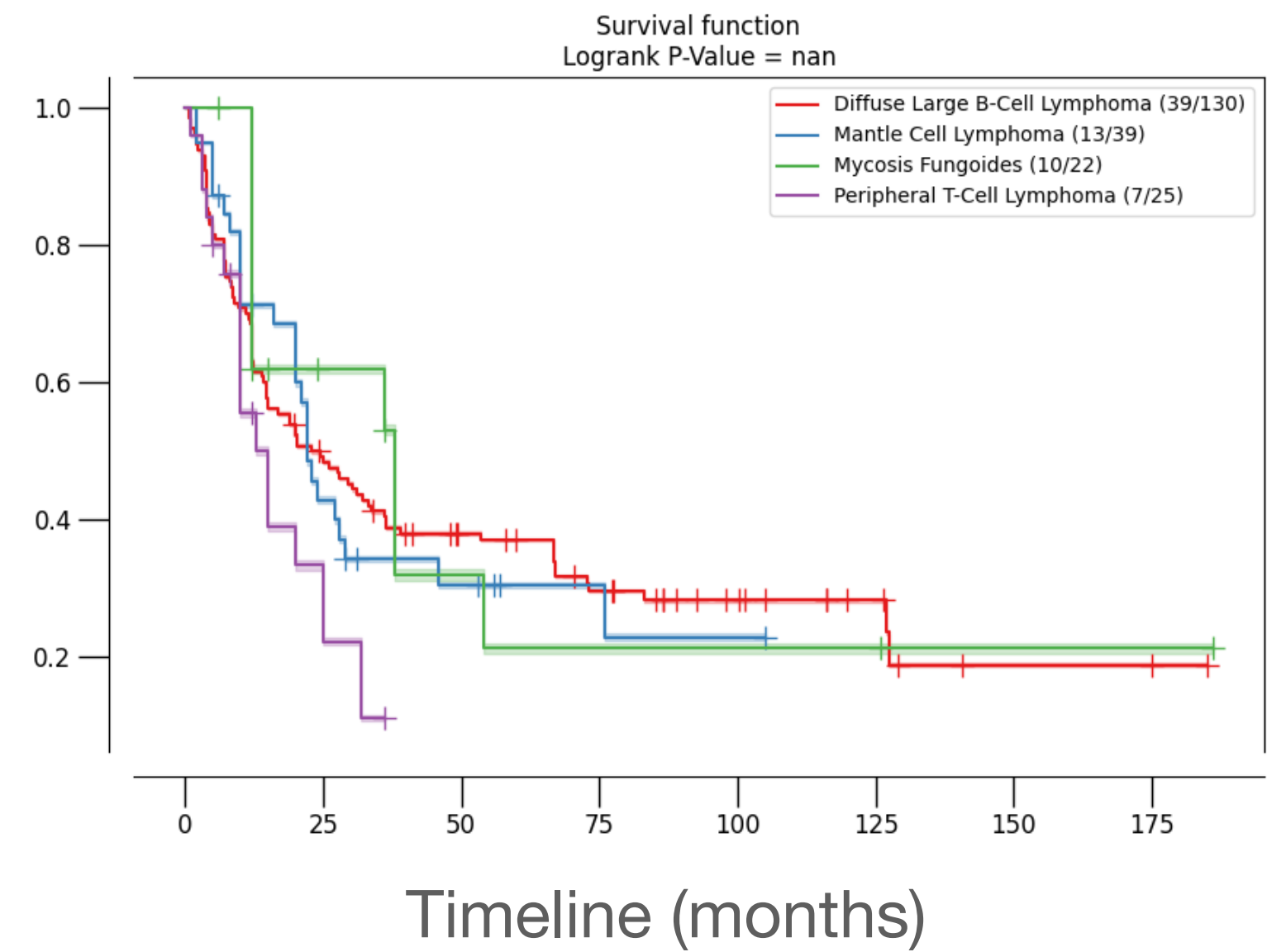
Results

KM Curves

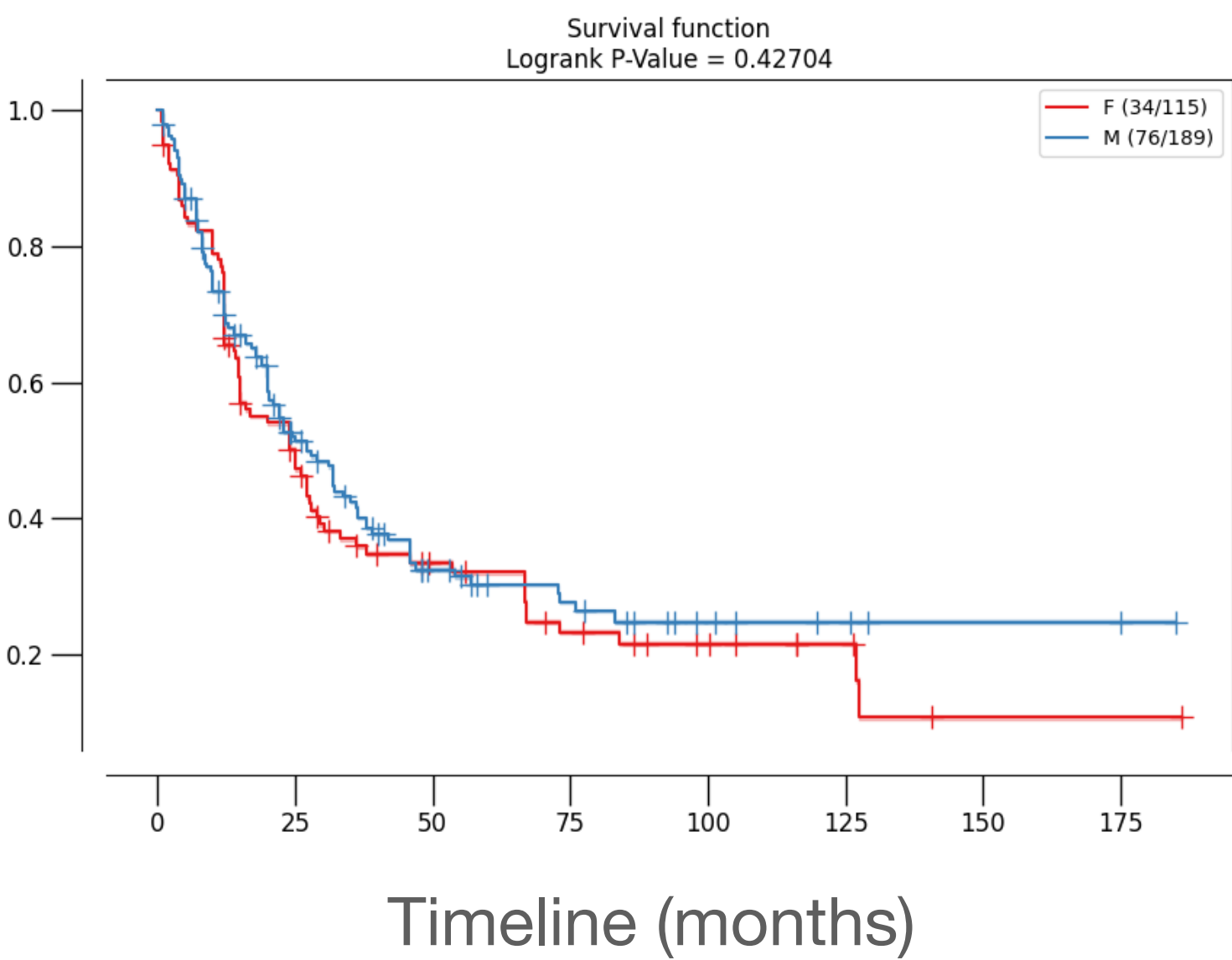
Survival rate based on gene



Survival rate based on tumor type

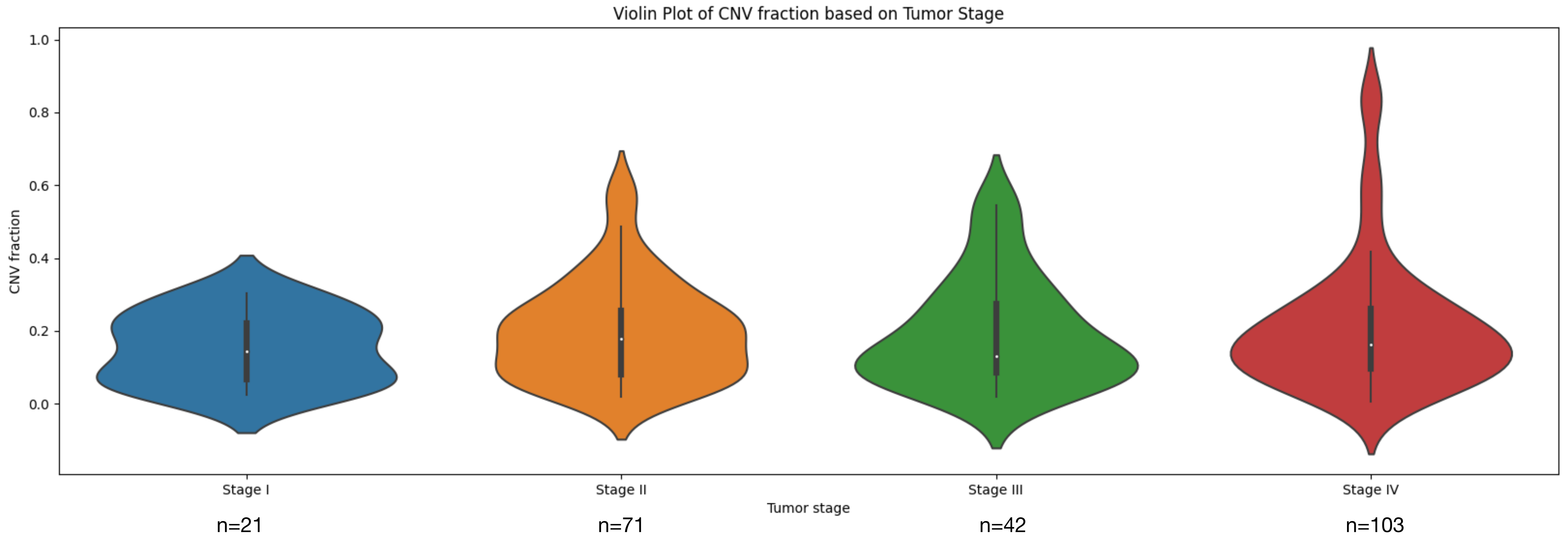


Survival rate based on sex of patient



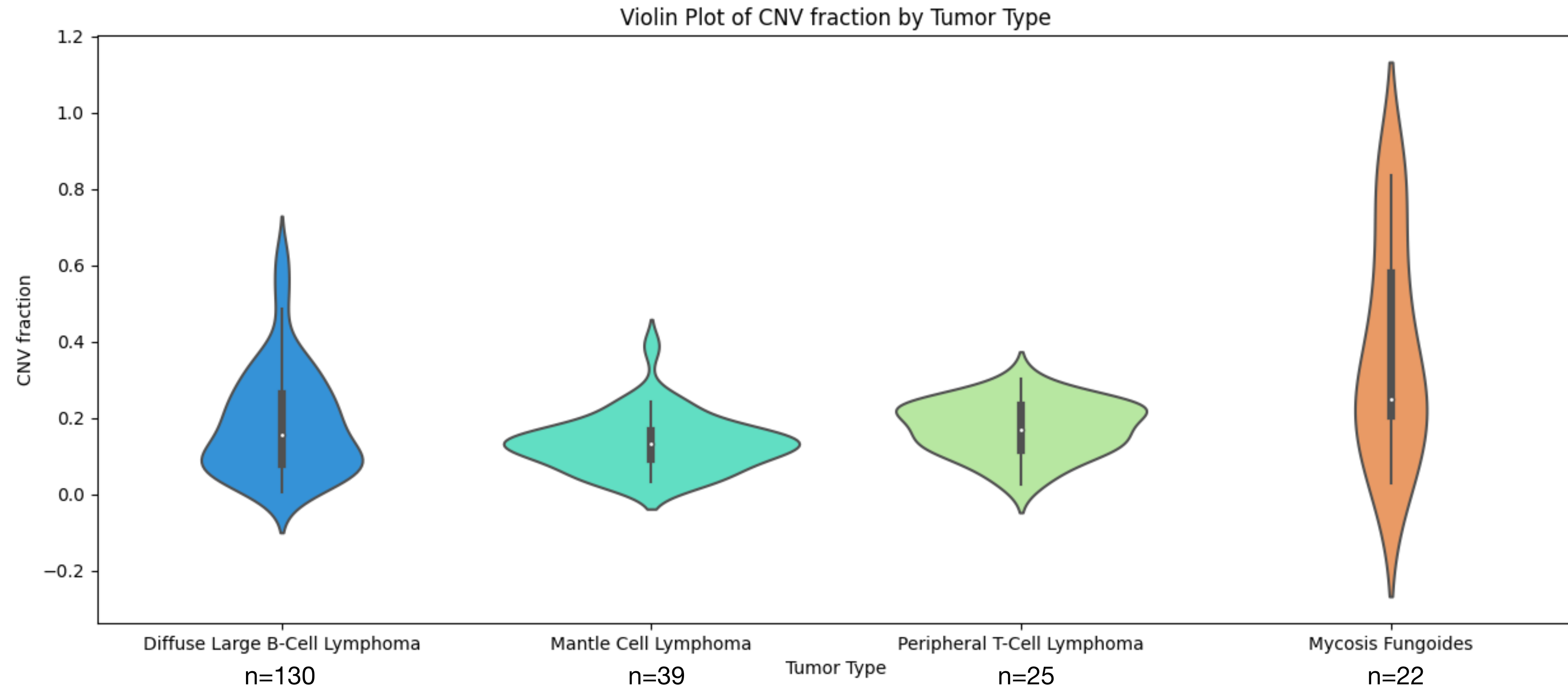
CNV fraction

Does CNV fraction increase with tumor stage?



CNV fraction

Does CNV fraction change across tumor types?



CNV fraction

Does CNV fraction change across samples with different mutated genes?

