University of Zurich[UZH]

BIO392
Bioinformatics of Genome Variations

Survival | Classifications

Michael Baudis **UZH SIB**
Computational Oncogenomics

BIO392 HS 2023

Github Activity

September 3, 2023 – October 3, 2023

Period: 1 month

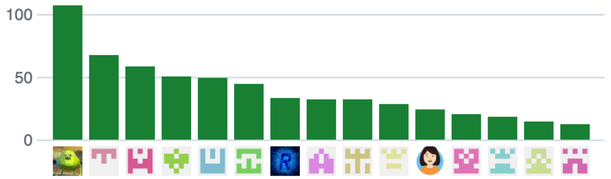
Overview

2 Active pull requests

0 Active issues

2 Merged pull requests	**0** Open pull requests	**0** Closed issues	**0** New issues

Excluding merges, **19 authors** have pushed **610 commits** to master and **611 commits** to all branches. On master, **299 files** have changed and there have been **10,596 additions** and **8,144 deletions**.



2 Pull requests merged by **1** person

- **Simplify wiggle**
  #220 merged 2 weeks ago
- **Update 2023 slides, copy exercises from 2022**
  #219 merged 2 weeks ago

BIO392 HS 2022

Github Activity

September 3, 2023 – October 3, 2023

Period: 1 month

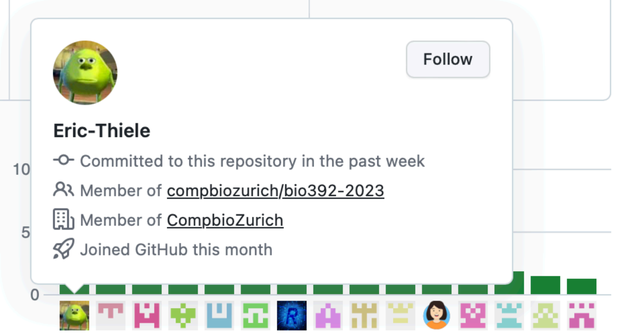
Overview

2 Active pull requests

0 Active issues

2
Merged pull requests

0
Open pull requests



Excluding merges, **19 authors** have pushed **610 commits** to master and **611 commits** to all branches. On master, **299 files** have changed and there have been **10,596 additions** and **8,144 deletions**.

2 Pull requests merged by **1** person

Simplify wiggle
#220 merged 2 weeks ago

Update 2023 slides, copy exercises from 2022
#219 merged 2 weeks ago

Survival

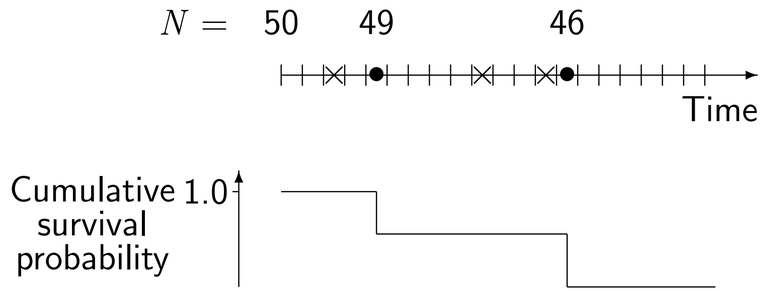
Kaplan-Meier Analysis of Survival Based on Conditional Probabilities

The Kaplan-Meier Method

- The most common method of estimating the survival function.
- A non-parametric method.
- Divides time into small intervals where the intervals are defined by the unique times of failure (death).
- Based on conditional probabilities as we are interested in the probability a subject surviving the next time interval given that they have survived so far.

Kaplan–Meier method illustrated

($\bullet$ = failure and $\times$ = censored):



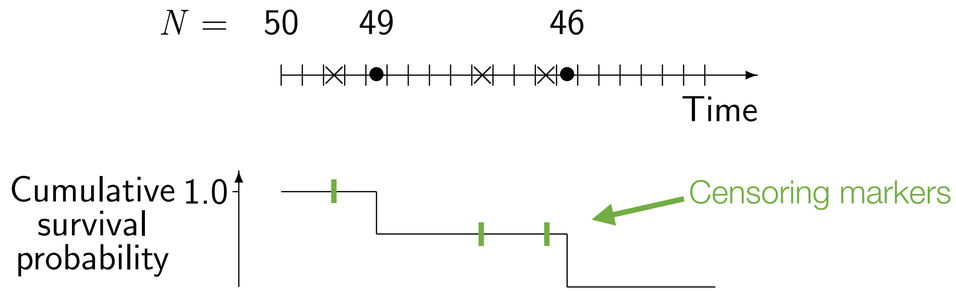
- Steps caused by multiplying by $(1 - 1/49)$ and $(1 - 1/46)$ respectively
- Late entry can also be dealt with

The Kaplan-Meier Method

- The most common method of estimating the survival function.
- A non-parametric method.
- Divides time into small intervals where the intervals are defined by the unique times of failure (death).
- Based on conditional probabilities as we are interested in the probability a subject surviving the next time interval given that they have survived so far.

Kaplan–Meier method illustrated

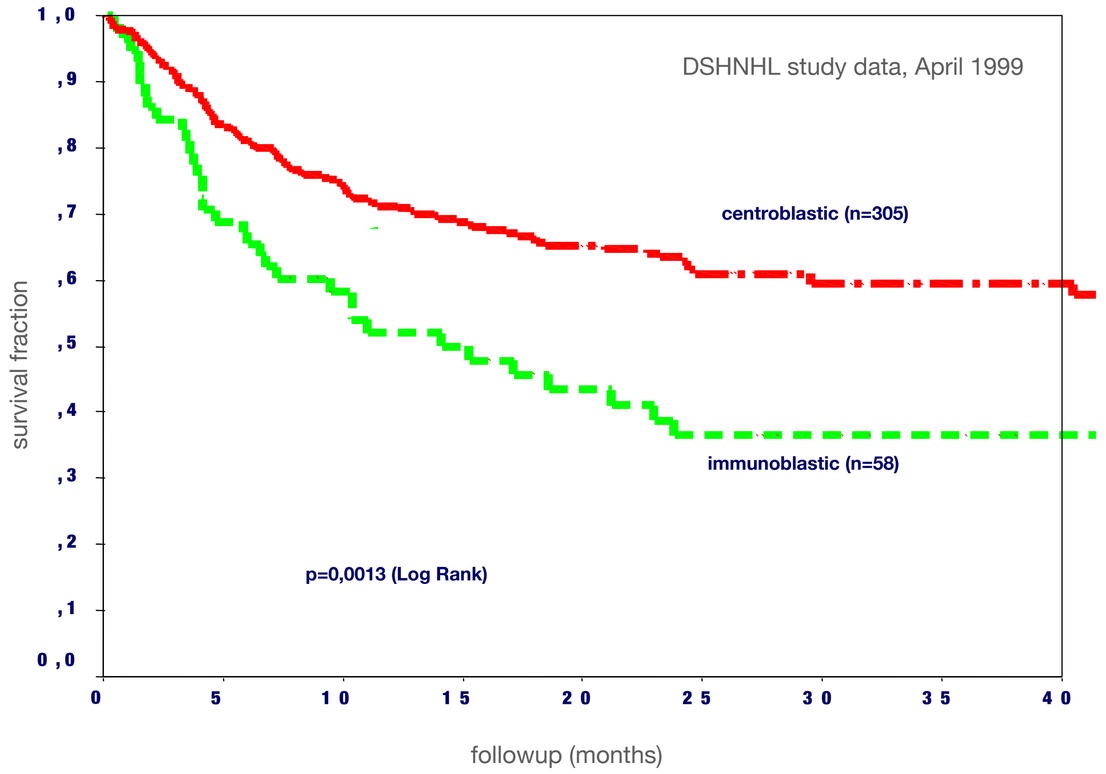
($\bullet$ = failure and $\times$ = censored):

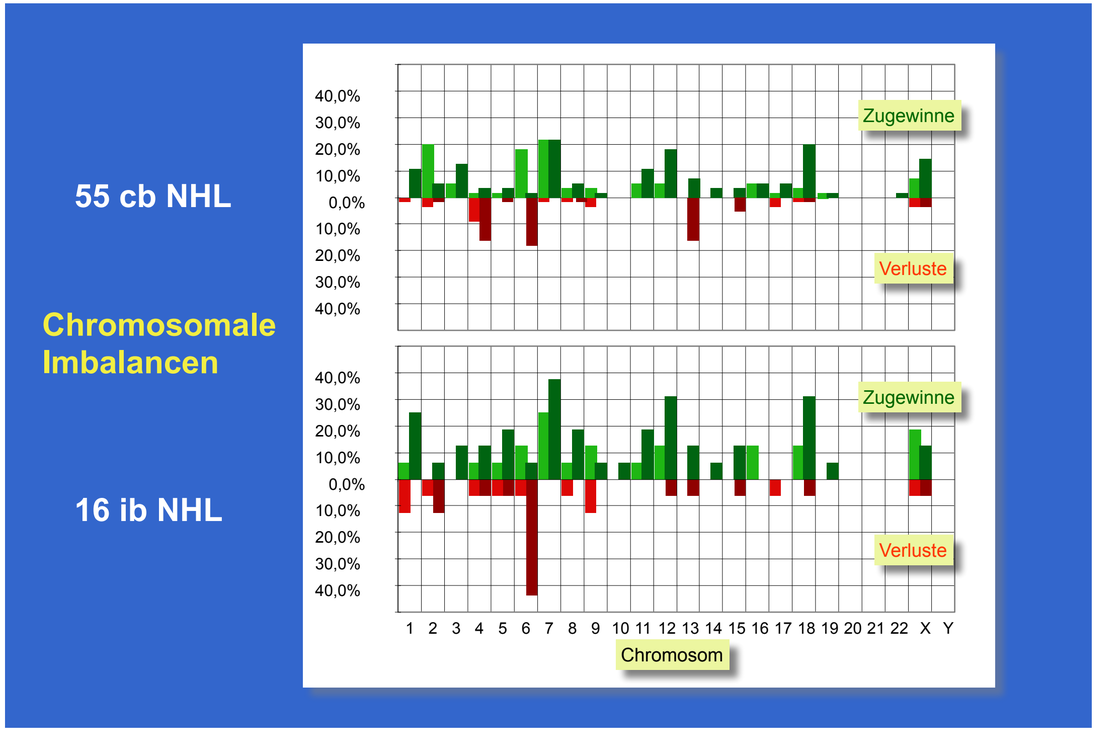


- Steps caused by multiplying by $(1 - 1/49)$ and $(1 - 1/46)$ respectively
- Late entry can also be dealt with

Cancer CNVs | Diagnostics | Prognosis

Single-study CNV frequencies correspond to diagnostic subsets





Kaplan-Meier Plots to Visualize Differential Risk

Multi-parametric "risk scores" in CLL Prognosis

Leukemia (2020) 34:1038–1051
https://doi.org/10.1038/s41375-020-0727-y

ARTICLE

Chronic lymphocytic leukemia

Prognostic model for newly diagnosed CLL patients in Binet stage A: results of the multicenter, prospective CLL1 trial of the German CLL study group

Manuela A. Hoechstetter[1] · Raymonde Busch[2] · Barbara Eichhorst[3] · Andreas Bühler[4] · Dirk Winkler[4] · Jasmin Bahlo[3] · Sandra Robrecht[3] · Michael J. Eckart[5] · Ursula Vehling-Kaiser[6] · Georg Jacobs[7] · Ulrich Jäger[8] · Hans Jürgen Hurtz[9] · Georg Hopfinger[10] · Frank Hartmann[11] · Harald Fuss[12] · Wolfgang Abenhardt[13] · Ilona Blau[14] · Werner Freier[15] · Lothar Müller[16] · Maria Goebeler[17] · Clemens Wendtner[1,3] · Kirsten Fischer[3] · Carmen D. Herling[3] · Michael Starck[1] · Martin Bentz[18] · Bertold Emmerich[19] · Hartmut Döhner[20] · Stephan Stilgenbauer[20] · Michael Hallek[3]

Table 2a Results of the Cox's regression for OS and TTFT in CLL patients in whom all 30 baseline parameters were available.

	Univariate comparison	Hazard ratio [HR]	95% Confidence Interval		*P* value
			Lower	Upper	
COX regression OS					
Cytogenetic Hierarchical Type					
del(17p)	**vs.** not del(17p)/del(11q)	3.8	2.1	7.1	<0.001
del(11q)	**vs.** not del(17p)/del(11q)	2.0	1.2	3.5	0.008
LDT					
<12 months	**vs.** ≥12 months	1.9	1.3	2.8	0.001
Age, years					
>60	**vs.** ≤60	1.8	1.2	2.7	0.002
B2M, mg/dL					
>3.5	**vs.** ≤3.5	2.0	1.2	3.1	0.004
IGHV mutational status					
Unmutated	**vs.** mutated	2.4	1.6	3.6	<0.001
COX regression TTFT					
Cytogenetic Hierarchical Type					
del(17p)	**vs.** not del(17p)/del(11q)	2.2	1.2	4.1	0.009
del(11q)	**vs.** not del(17p)/del(11q)	2.0	1.3	3.0	0.001
LDT	**vs.**	2.3	1.7	3.1	<0.001
Age, years					
>60	**vs.** ≤60	1.3	1.0	1.7	0.037
B2M, mg/dL					
>3.5	**vs.** ≤3.5	1.5	1.0	2.3	0.049
IGHV mutational status					
Unmutated	**vs.** mutated	4.4	3.2	5.9	<0.001

Table 2b Allocation of risk score points to the distinctive factors of the CLL1-PM.

	HR (95% CI)	*P*	Allocated risk score points
Characteristics			
Del(17p)	3.8 (2.1–7.1)	<0.001	3.5
Unmutated *IGHV*	2.4 (1.6–3.6)	<0.001	2.5
Del(11q)	2.0 (1.2–3.5)	0.008	2.5
Beta2-MG >3.5 mg/L	2.0 (1.2–3.1)	0.004	2.5
LDT<12 months	1.9 (1.3–2.8)	0.001	1.5
Age >60 years	1.8 (1.2–2.7)	0.002	1.5

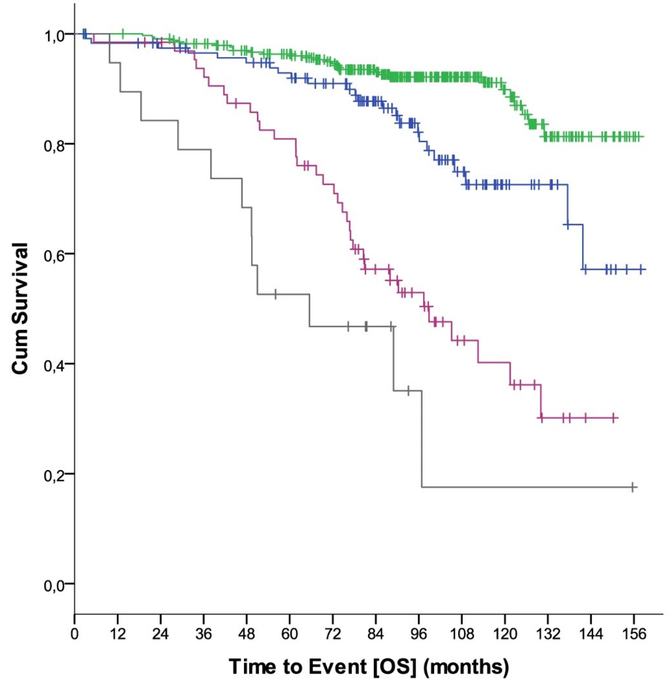
The assigned risk score points derived from the HR for OS of the individual factors.

Table 2c Patients and risk groups according to the CLL1 Prognostic Model (CLL1-PM). Patients and risk groups according to the CLL-IPI.

	Index score	Patients *N* (%)
Risk Groups accroding to the CLL1-PM		539
Very low	0.0–1.5	336 (62.3)
Low	2.0–4.0	119 (22.1)
High	4.5–6.5	65 (12.1)
Very high	7.0–14.0	19 (3.5)
Risk Groups according to the CLL-IPI		539
Low	0–1	360 (66.8)
Intermediate	2–3	141 (26.2)
High	4–6	33 (6.1)
Very high	7–10	5 (0.9)

OS overall survival, *HR* hazard ratio, *Beta2-MG* beta-2 microglobulin, *IGHV* immunoglobulin heavy-chain genes, *LDT* lymphocyte doubling time, *TTFT* time-to-first treatment.

- "a novel prognostic model (CLL1-PM) developed to identify risk groups, separating patients with favorable from others with dismal prognosis"
- " findings would be useful to effectively stratify Binet stage A patients, particularly within the scope of clinical trials evaluating novel agents"



***P* < 0.001**

Number at risk	0	12	24	36	48	60	72	84	96	108	120	132	144	156
Very low	336	335	331	322	306	294	262	215	160	113	68	33	15	2
Low	119	115	111	108	106	100	89	71	49	34	19	14	6	1
High	65	64	63	59	54	50	43	29	21	12	10	4	1	0
Very high	19	18	16	15	13	9	8	5	2	1	1	1	1	0

Discrimination: ***C*-statistics, *C* = 0.739 (95% CI, 0.686– 0.790)**
AIC=445

Overall survival according to the CLL1-PM risk groups. The full analysis dataset is comprised of the dataset of 539 patients.

2018-12-28___Yosifov,-Stilgenbauer-et-al.__The-CLL-Success-Story__HemaSphere-Review.pdf

Kaplan-Meier Plots to Visualize Differential Risk

Multi-parametric "risk scores" in CLL Prognosis

Leukemia (2020) 34:1038–1051
https://doi.org/10.1038/s41375-020-0727-y

ARTICLE

Chronic lymphocytic leukemia

Prognostic model for newly diagnosed CLL patients in Binet stage A: results of the multicenter, prospective CLL1 trial of the German CLL study group

Manuela A. Hoechstetter[1] · Raymonde Busch[2] · Barbara Eichhorst[3] · Andreas Bühler[4] · Dirk Winkler[4] · Jasmin Bahlo[3] · Sandra Robrecht[3] · Michael J. Eckart[5] · Ursula Vehling-Kaiser[6] · Georg Jacobs[7] · Ulrich Jäger[8] · Hans Jürgen Hurtz[9] · Georg Hopfinger[10] · Frank Hartmann[11] · Harald Fuss[12] · Wolfgang Abenhardt[13] · Ilona Blau[14] · Werner Freier[15] · Lothar Müller[16] · Maria Goebeler[17] · Clemens Wendtner[1,3] · Kirsten Fischer[3] · Carmen D. Herling[3] · Michael Starck[1] · Martin Bentz[18] · Bertold Emmerich[19] · Hartmut Döhner[20] · Stephan Stilgenbauer[20] · Michael Hallek[3]

Table 2a Results of the Cox's regression for OS and TTFT in CLL patients in whom all 30 baseline parameters were available.

	Univariate comparison	Hazard ratio [HR]	95% Confidence Interval		*P* value
			Lower	Upper	
COX regression OS					
Cytogenetic Hierarchical Type					
del(17p)	**vs.** not del(17p)/del(11q)	3.8	2.1	7.1	<0.001
del(11q)	**vs.** not del(17p)/del(11q)	2.0	1.2	3.5	0.008
LDT					
<12 months	**vs.** ≥12 months	1.9	1.3	2.8	0.001
Age, years					
>60	**vs.** ≤60	1.8	1.2	2.7	0.002
B2M, mg/dL					
>3.5	**vs.** ≤3.5	2.0	1.2	3.1	0.004
IGHV mutational status					
Unmutated	**vs.** mutated	2.4	1.6	3.6	<0.001
COX regression TTFT					
Cytogenetic Hierarchical Type					
del(17p)	**vs.** not del(17p)/del(11q)	2.2	1.2	4.1	0.009
del(11q)	**vs.** not del(17p)/del(11q)	2.0	1.3	3.0	0.001
LDT	**vs.**	2.3	1.7	3.1	<0.001
Age, years					
>60	**vs.** ≤60	1.3	1.0	1.7	0.037
B2M, mg/dL					
>3.5	**vs.** ≤3.5	1.5	1.0	2.3	0.049
IGHV mutational status					
Unmutated	**vs.** mutated	4.4	3.2	5.9	<0.001

Table 2b Allocation of risk score points to the distinctive factors of the CLL1-PM.

	HR (95% CI)	*P*	Allocated risk score points
Characteristics			
Del(17p)	3.8 (2.1–7.1)	<0.001	3.5
[illegible]	[illegible]	[illegible]	[illegible]
Del(11q)	2.0 (1.2–3.5)	0.008	2.5
Beta2-MG >3.5 mg/L	2.0 (1.2–3.1)	0.004	2.5
LDT<12 months	1.9 (1.3–2.8)	0.001	1.5
Age >60 years	1.8 (1.2–2.7)	0.002	1.5

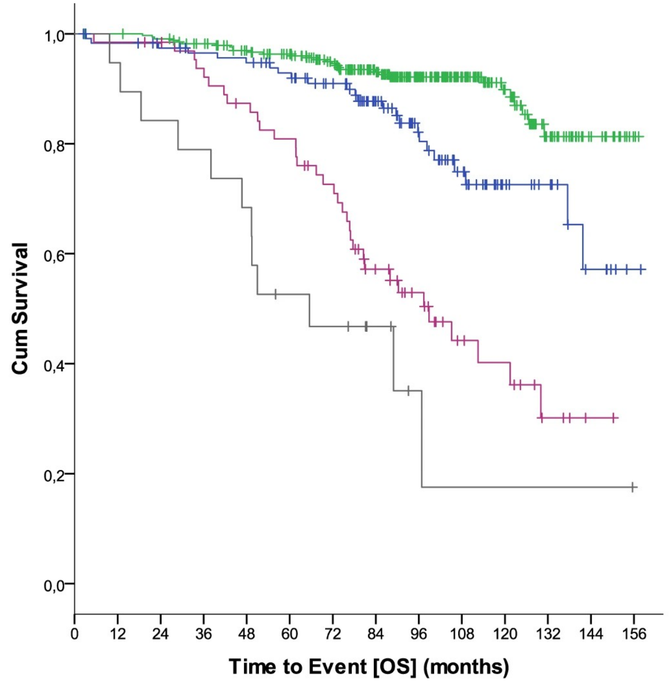
The assigned risk score points derived from the HR for OS of the individual factors.

Table 2c Patients and risk groups according to the CLL1 Prognostic Model (CLL1-PM). Patients and risk groups according to the CLL-IPI.

	Index score	Patients *N* (%)
Risk Groups accroding to the CLL1-PM		539
Very low	0.0–1.5	336 (62.3)
Low	2.0–4.0	119 (22.1)
High	4.5–6.5	65 (12.1)
Very high	7.0–14.0	19 (3.5)
Risk Groups according to the CLL-IPI		539
Low	0–1	360 (66.8)
Intermediate	2–3	141 (26.2)
High	4–6	33 (6.1)
Very high	7–10	5 (0.9)

OS overall survival, *HR* hazard ratio, *Beta2-MG* beta-2 microglobulin, *IGHV* immunoglobulin heavy-chain genes, *LDT* lymphocyte doubling time, *TTFT* time-to-first treatment.

- "a novel prognostic model (CLL1-PM) developed to identify risk groups, separating patients with favorable from others with dismal prognosis"
- " findings would be useful to effectively stratify Binet stage A patients, particularly within the scope of clinical trials evaluating novel agents"



***P* < 0.001**

Number at risk	0	12	24	36	48	60	72	84	96	108	120	132	144	156
Very low	336	335	331	322	306	294	262	215	160	113	68	33	15	2
Low	119	115	111	108	106	100	89	71	49	34	19	14	6	1
High	65	64	63	59	54	50	43	29	21	12	10	4	1	0
Very high	19	18	16	15	13	9	8	5	2	1	1	1	1	0

Discrimination: ***C*-statistics, *C* = 0.739 (95% CI, 0.686– 0.790)**
AIC=445

Overall survival according to the CLL1-PM risk groups. The full analysis dataset is comprised of the dataset of 539 patients.

2018-12-28___Yosifov,-Stilgenbauer-et-al.__The-CLL-Success-Story__HemaSphere-Review.pdf

Cancer Classifications & Parameters

NCIt | ICD-O / WHO | TNM

International Agency for Research on Cancer
World Health Organization

ICD-O 3

WHO International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3)

- used in cancer registries for coding the site (topography) and the histology (morphology) of neoplasms, usually obtained from a pathology report
- mix of "biology" (i.e. tumor morphology) and "clinical" (i.e. tumor site)

  ➡ 2 codes per cancer

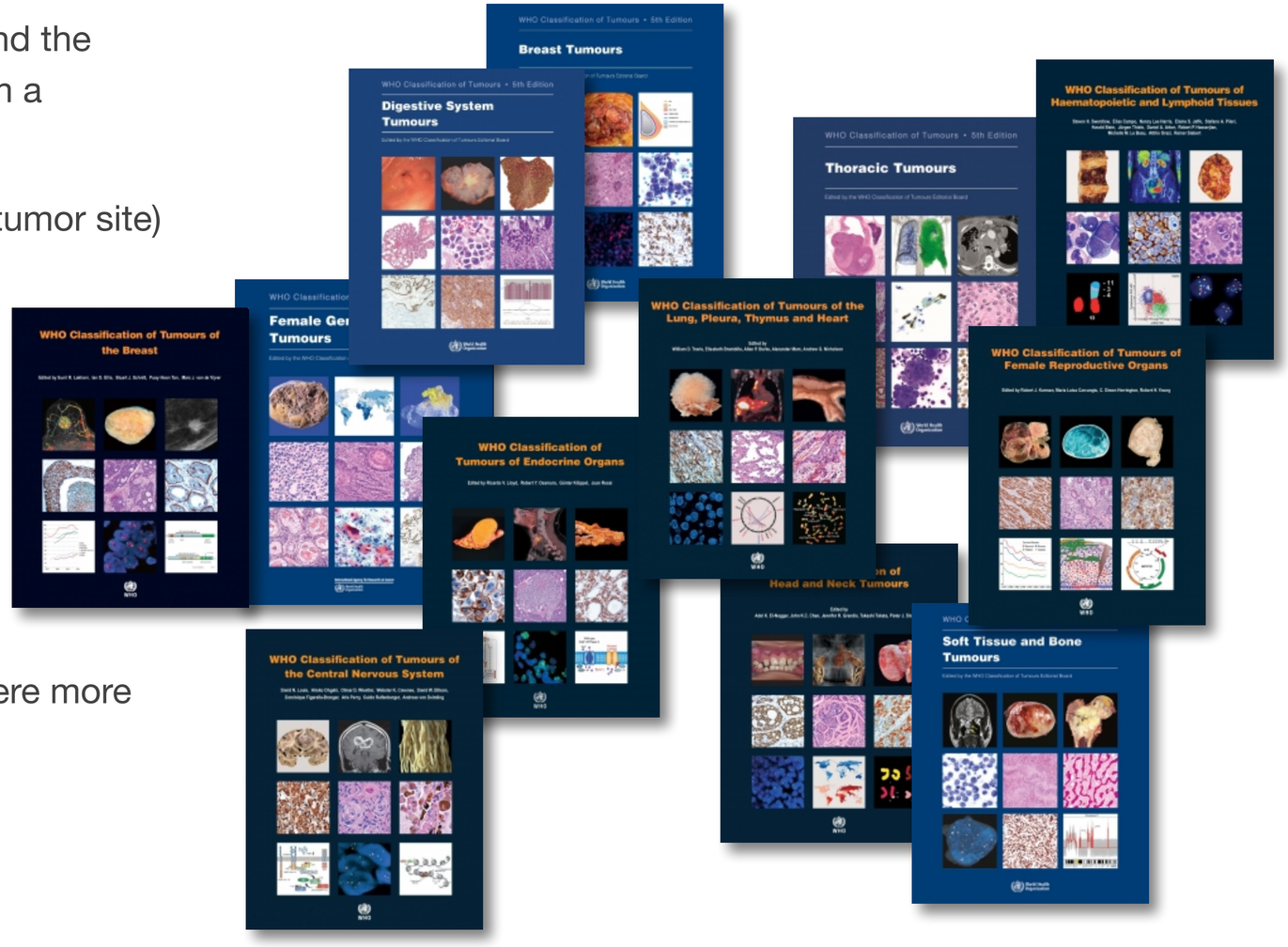
  - "Adenocarcinoma" of the "Sigmoid colon"

    8140/3 **C18.7**

  - "Retinoblastoma" of the "Retina"

    9510/3 **C69.2**

- widely accepted by pathologists but limited clinical use (there more ICD-10 or SNOMED)
- no ontology & not (truly) hierarchical
- many entities difficult to remap if using only single code



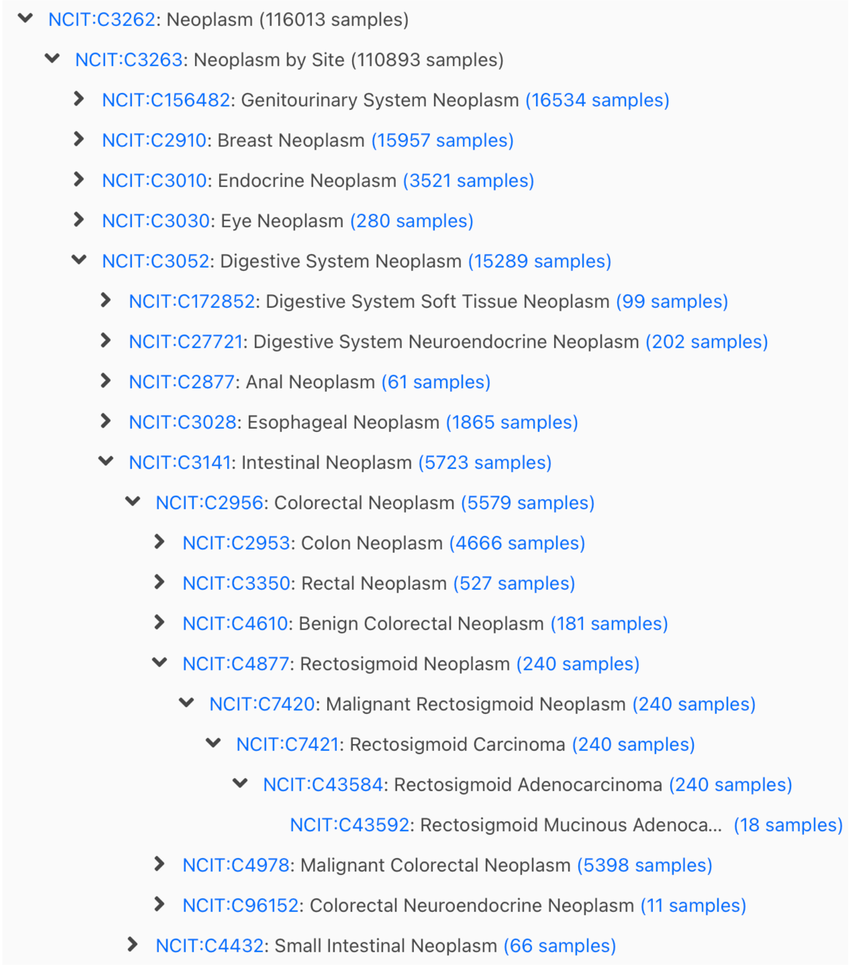
NCIt

Neoplasm Classifications in the NCI Thesaurus

- NCI's core reference terminology and biomedical ontology are collected in the NCI Thesaurus (NCIt)
- individual codes for site-specific occurrences of "biological" diagnoses

  1 code per cancer

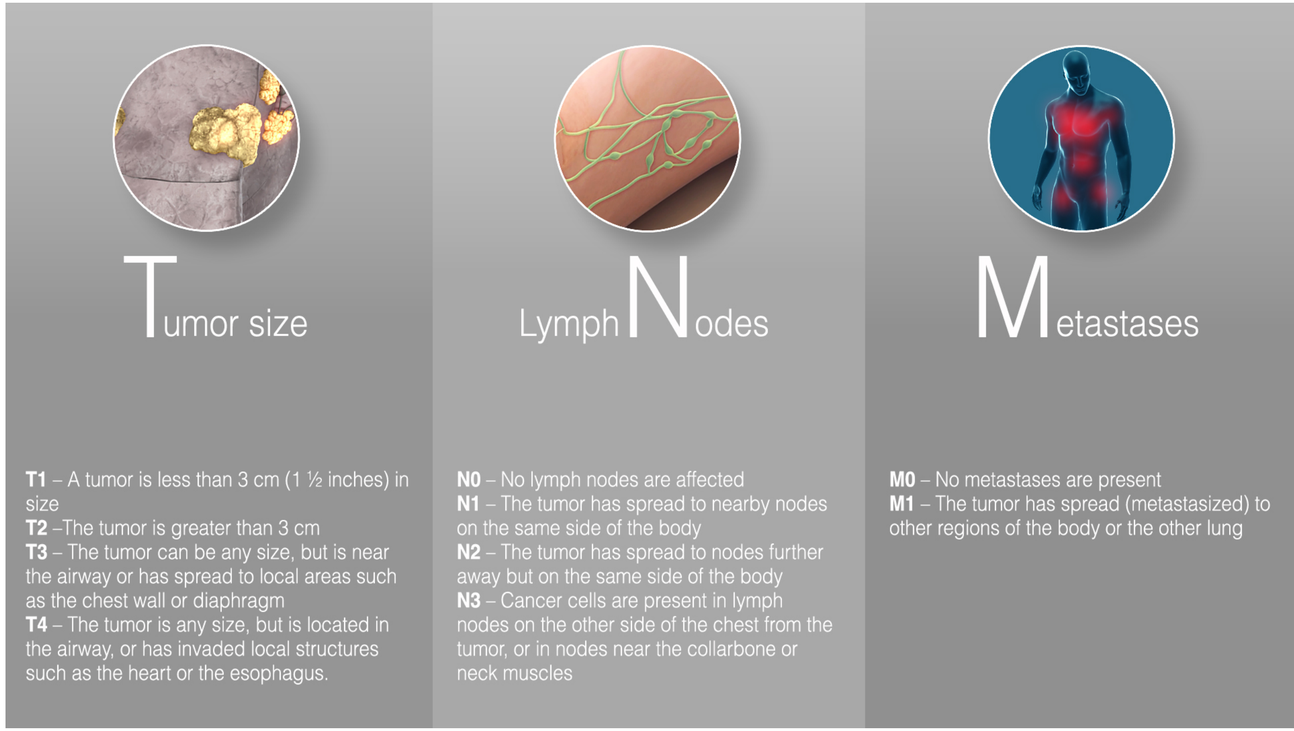
  - **NCIT:C43584** - Rectosigmoid Adenocarcinoma
  - **NCIT:C7541** - Retinoblastoma
- truly hierarchical ontology
- hierarchical system empowers "logical OR" queries
- terms can have multiple occurrences in diagnostic tree
- assignment of code to different groupings allows soft aggregation (e.g. a type of colorectal adenocarcinoma with all colon tumors or with all adenocarcinomas)



TNM

A Classification for Clinical Cancer Stage Parameters

- most widely used cancer staging system
- **T** refers to the size and extent of the main tumor
- **N** refers to the the number / location of nearby lymph nodes that have cancer infiltration
- **M** refers to whether the cancer has metastasized
- not used for leukemias / lymphomas
  - Binet and Rai in CLL
  - proportion of blasts in bone marrow or blood in leukemias
  - Lugano classification in lymphomas
- other disease specific staging systems may (co-) exist
  - e.g. a stage II breast cancer is determined by size & nodal involvement



Source: www.scientificanimations.com

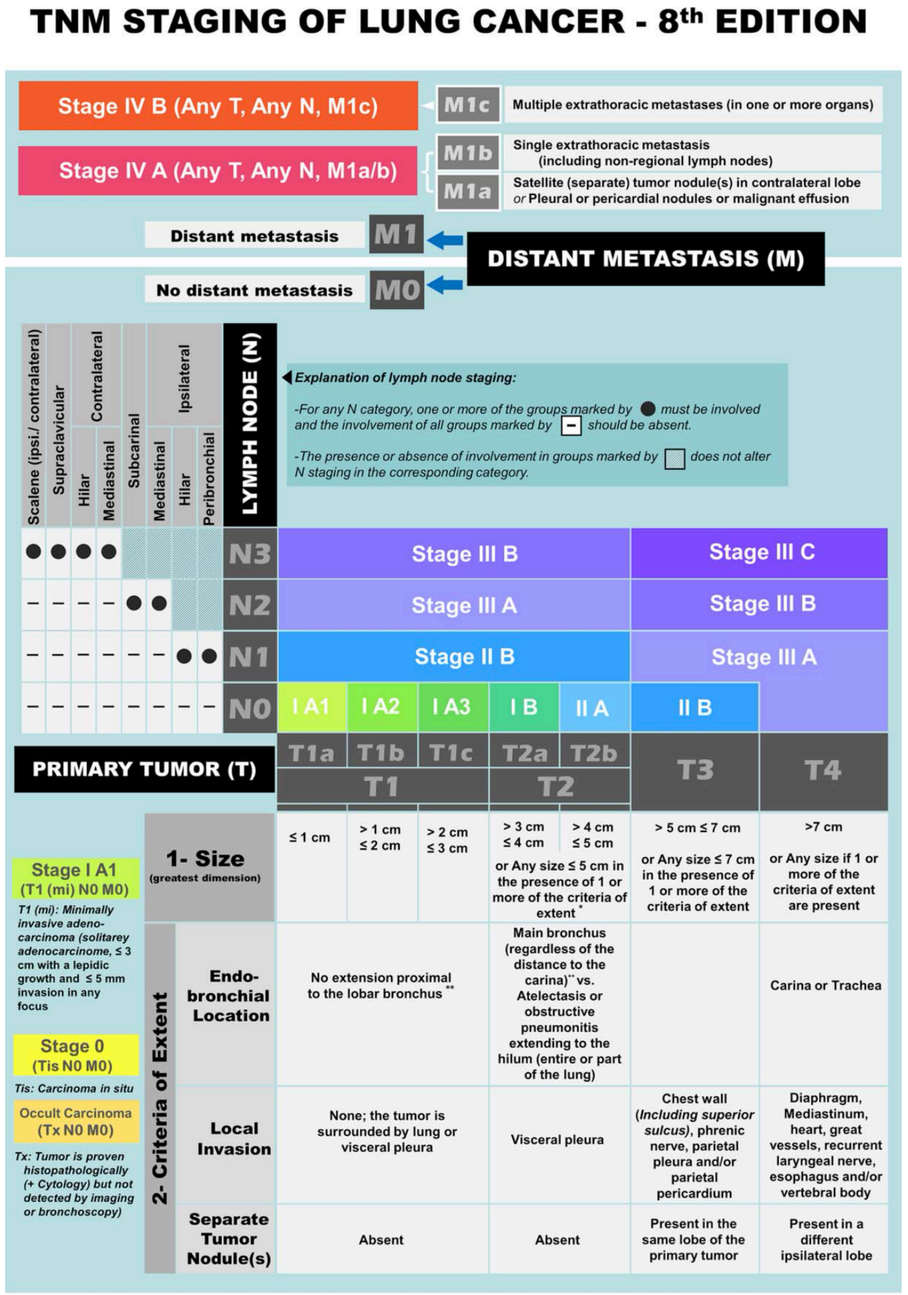
Source: https://www.cancer.gov/about-cancer/diagnosis-staging/staging

TNM

A Classification for Clinical Cancer Stage Parameters

- most widely used cancer staging system
- **T** refers to the size and extent of the main tumor
- **N** refers to the the number / location of nearby lymph nodes that have cancer infiltration
- **M** refers to whether the cancer has metastasized
- not used for leukemias / lymphomas
  - Binet and Rai in CLL
  - proportion of blasts in bone marrow or blood in leukemias
  - Lugano classification in lymphomas
- other disease specific staging systems may (co-) exist
  - e.g. a stage II lung cancer is determined by size & nodal involvement

Source: https://www.cancer.gov/about-cancer/diagnosis-staging/staging



Lababede O, Meziane MA. The Eighth Edition of TNM Staging of Lung Cancer: Reference Chart and Diagrams. Oncologist. 2018;23(7):844-848.

TNM

A Classification for Clinical Cancer Stage Parameters

- most widely used cancer staging system
- **T** refers to the size and extent of the main tumor
- **N** refers to the the number / location of nearby lymph nodes that have cancer infiltration
- **M** refers to whether the cancer has metastasized
- not used for leukemias / lymphomas
  - Binet and Rai in CLL
  - proportion of blasts in bone marrow or blood in leukemias
  - Lugano classification in lymphomas
- other disease specific staging systems may (co-) exist
  - e.g. a stage II lung cancer is determined by size & nodal involvement

TNM has been "ontologized" into NCIt

Source: https://www.cancer.gov/about-cancer/diagnosis-staging/staging

NCIT:C48698	Cancer TNM Finding Category	0
NCIT:C133398	Postneoadjuvant Therapy Pathologic TNM Finding	1
NCIT:C143081	Posttherapy Clinical TNM Finding	1
NCIT:C48739	Pathologic TNM Finding	1
NCIT:C48886	Pathologic Distant Metastasis TNM Finding	2
NCIT:C48740	pM0 Stage Finding	3
NCIT:C48741	pM1 Stage Finding	3
NCIT:C48742	pM1a Stage Finding	4
NCIT:C48743	pM1b Stage Finding	4
NCIT:C48744	pM1c Stage Finding	4
NCIT:C48887	Pathologic Regional Lymph Nodes TNM Finding	2
NCIT:C48745	pN0 Stage Finding	3
NCIT:C48746	pN1 Stage Finding	3
NCIT:C48747	pN1a Stage Finding	4
NCIT:C48748	pN1b Stage Finding	4
NCIT:C48749	pN1c Stage Finding	4
NCIT:C48750	pN2 Stage Finding	3
NCIT:C48751	pN2a Stage Finding	4
NCIT:C48752	pN2b Stage Finding	4
NCIT:C48753	pN2c Stage Finding	4
NCIT:C48754	pN3 Stage Finding	3
NCIT:C48755	pN3a Stage Finding	4
NCIT:C48756	pN3b Stage Finding	4
NCIT:C48757	pN3c Stage Finding	4
NCIT:C48888	Pathologic Primary Tumor TNM Finding	2
NCIT:C48758	pT0 Stage Finding	3
NCIT:C48759	pT1 Stage Finding	3
NCIT:C48760	pT1a Stage Finding	4
NCIT:C48761	pT1b Stage Finding	4
NCIT:C48763	pT1c Stage Finding	4
NCIT:C48764	pT2 Stage Finding	3
NCIT:C48765	pT2a Stage Finding	4
NCIT:C48766	pT2b Stage Finding	4
NCIT:C48767	pT2c Stage Finding	4
NCIT:C48768	pT3 Stage Finding	3
NCIT:C48769	pT3a Stage Finding	4
NCIT:C48770	pT3b Stage Finding	4
NCIT:C48771	pT3c Stage Finding	4
NCIT:C48772	pT4 Stage Finding	3
NCIT:C48773	pT4a Stage Finding	4
NCIT:C48774	pT4b Stage Finding	4
NCIT:C48775	pT4c Stage Finding	4
NCIT:C48776	pT4d Stage Finding	4
NCIT:C48879	Generic TNM Finding	1
NCIT:C48777	Cancer TNM Vessel Invasion Finding Category	2
NCIT:C147091	Lymphovascular Invasion 0	3
NCIT:C147092	Lymphovascular Invasion 1	3
NCIT:C147093	Lymphovascular Invasion 9	3
NCIT:C147094	Lymphovascular Invasion 2	3
NCIT:C147095	Lymphovascular Invasion 3	3
NCIT:C147096	Lymphovascular Invasion 4	3
NCIT:C48883	Generic Distant Metastasis TNM Finding	2
NCIT:C48699	M0 Stage Finding	3
NCIT:C95956	cM0 (i+) Stage Finding	4
NCIT:C48700	M1 Stage Finding	3
NCIT:C48701	M1a Stage Finding	4
NCIT:C48702	M1b Stage Finding	4
NCIT:C48703	M1c Stage Finding	4

NCIT:C48704	MX Stage Finding	0
NCIT:C48884	Generic Regional Lymph Nodes TNM Finding	1
NCIT:C48705	N0 Stage Finding	1
NCIT:C95921	N0 (i-) Stage Finding	1
NCIT:C95922	N0 (i+) Stage Finding	2
NCIT:C95923	N0 (mol-) Stage Finding	3
NCIT:C95925	N0 (mol+) Stage Finding	3
NCIT:C48706	N1 Stage Finding	4
NCIT:C48707	N1a Stage Finding	4
NCIT:C48708	N1b Stage Finding	4
NCIT:C95929	N1bI Stage Finding	2
NCIT:C95935	N1bII Stage Finding	3
NCIT:C95936	N1bIII Stage Finding	3
NCIT:C95937	N1bIV Stage Finding	4
NCIT:C48709	N1c Stage Finding	4
NCIT:C95955	N1mi Stage Finding	4
NCIT:C48714	N3 Stage Finding	3
NCIT:C48715	N3a Stage Finding	4
NCIT:C48716	N3b Stage Finding	4
NCIT:C48717	N3c Stage Finding	4
NCIT:C48718	NX Stage Finding	3
NCIT:C48786	N2 Stage Finding	4
NCIT:C48711	N2a Stage Finding	4
NCIT:C48712	N2b Stage Finding	4
NCIT:C48713	N2c Stage Finding	2
NCIT:C96026	N4 Stage Finding	3
NCIT:C48885	Generic Primary Tumor TNM Finding	3
NCIT:C106299	Any T	4
NCIT:C132010	T5 Stage Finding	4
NCIT:C48719	T0 Stage Finding	4
NCIT:C48720	T1 Stage Finding	3
NCIT:C48721	T1a Stage Finding	4
NCIT:C48722	T1b Stage Finding	4
NCIT:C48723	T1c Stage Finding	4
NCIT:C95805	T1mi Stage Finding	3
NCIT:C48724	T2 Stage Finding	4
NCIT:C148411	T2d Stage Finding	4
NCIT:C48725	T2a Stage Finding	4
NCIT:C48726	T2b Stage Finding	3
NCIT:C48727	T2c Stage Finding	4
NCIT:C48728	T3 Stage Finding	4
NCIT:C148412	T3d Stage Finding	4
NCIT:C48729	T3a Stage Finding	4
NCIT:C48730	T3b Stage Finding	1
NCIT:C48731	T3c Stage Finding	2
NCIT:C48732	T4 Stage Finding	3
NCIT:C48733	T4a Stage Finding	3
NCIT:C48734	T4b Stage Finding	3
NCIT:C48735	T4c Stage Finding	3
NCIT:C48736	T4d Stage Finding	3
NCIT:C48737	TX Stage Finding	3
NCIT:C48738	Tis Stage Finding	2
NCIT:C96025	Ta Stage Finding	3
NCIT:C48880	Recurrent Cancer TNM Finding	4
NCIT:C48881	Clinical TNM Finding	3
NCIT:C161009	Clinical Primary Tumor TNM Finding	4
NCIT:C162609	Clinical Regional Lymph Nodes TNM Finding	4
NCIT:C162610	Clinical Distant Metastasis TNM Finding	4
NCIT:C48882	Autopsy TNM Finding	

Tasks

Survival analyses | Cancer classifications | Staging

- Familiarize yourself with the different concepts behind different disease clasification systems - what are there use, advantages, problems? E.g. ICD-10, ICD-O, NCIt
  - you can use Progenetix to explore e.g. ontology mapping
- Learn to "read" Kaplan-Meier plots (preparation for explorative analyses later this week).
- Achieve a principal understanding of TNM codes & write some "translations"

➡ T1N1M0: small tumor with regional lymph node involvement and no detected distant metastases