Driver Gene Identification

Federico Abascal
fa8@sanger.ac.uk
Martincorena's lab
Wellcome Sanger Institute

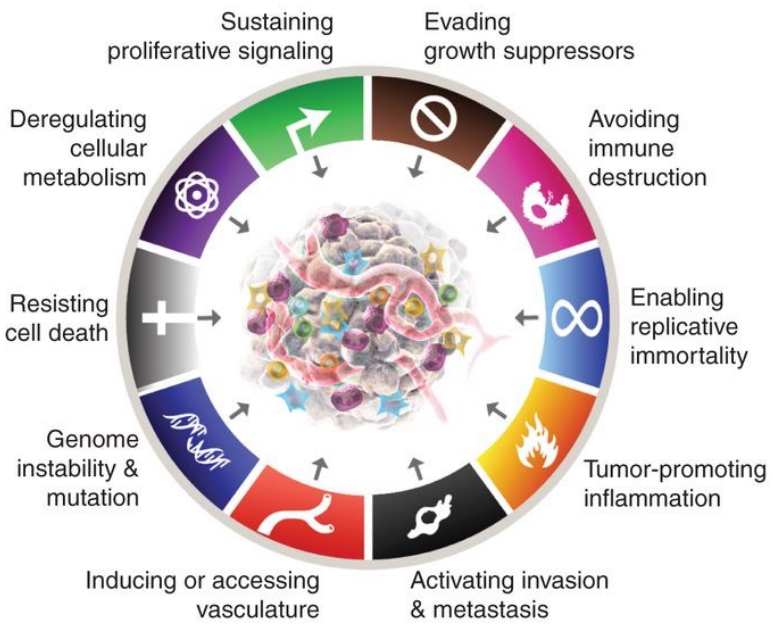
Day 3

- Drivers: theory and detection, *F. Abascal*
  - Exercises with `dndscv`
- Invited seminar: Cancer genomics in Latin America, *Daniela Robles*
- OncoPlots, *P. Basurto*
  - Introduction and exercises with `maftools`
- Discussion & Wrap-up

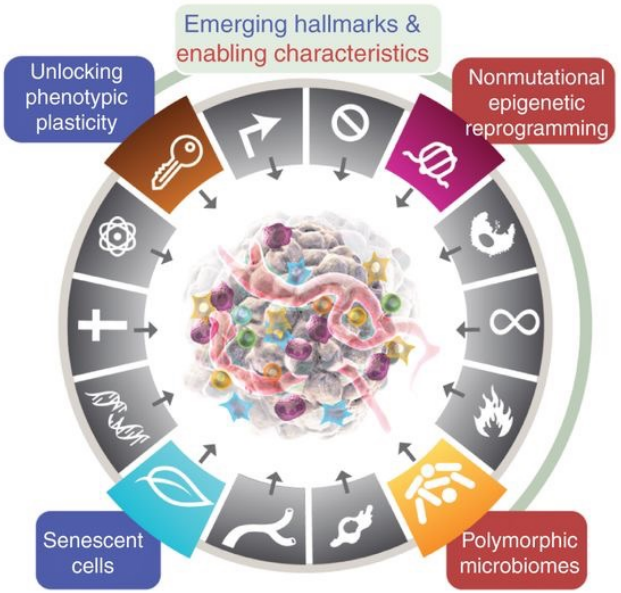
What are drivers and passengers?

- *American Journal of Traffic and Transportation Engineering*
- Drivers are **causative alterations**
  - Substitutions and small indels – point mutations:
    - Coding alterations: KRAS G12D
    - Regulatory region: *TERT* promoter
  - Structural rearrangements and copy number changes:
    - *BCR-ABL1* in leukemia, *MYC* amplification, long deletions (*TP53*)...
  - Epigenetic alterations:
    - *VHL* expression repression through promoter hypermethylation

Driver functions – hallmarks of cancer



Hanahan & Weinberg, 2011, *Cell*



Hanahan, 2022, *Cancer Discovery*

Cancer Gene Census
https://cancer.sanger.ac.uk/census
Lots of info!

Tier1: 579 genes
Tier2: 154 genes

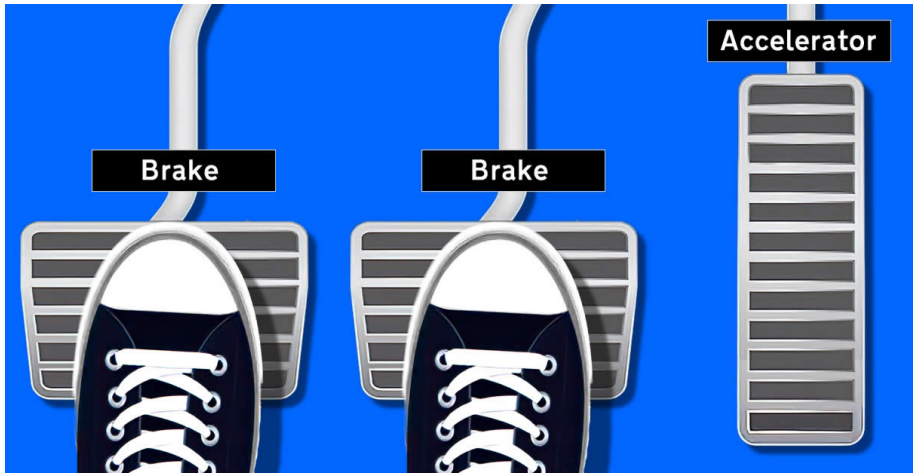
(WG < 20,000 protein-coding genes)

Oncogenes & tumour suppressors

Tumour suppressors **Oncogenes**

Loss of function

- Double hit
- Missense, splice site, nonsense mutations
- Deletions and insertions
- Loss of loci
  - *TP53*
  - *VHL*



Gain of function

- Single hit
- Missense mutations
- Copy number gain (amplification)
  - *TERT* promoter
  - *KRAS*
  - *MYC*

How many mutations are required to develop a tumour?

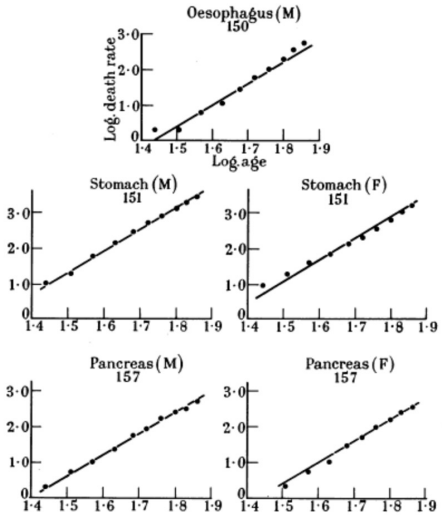
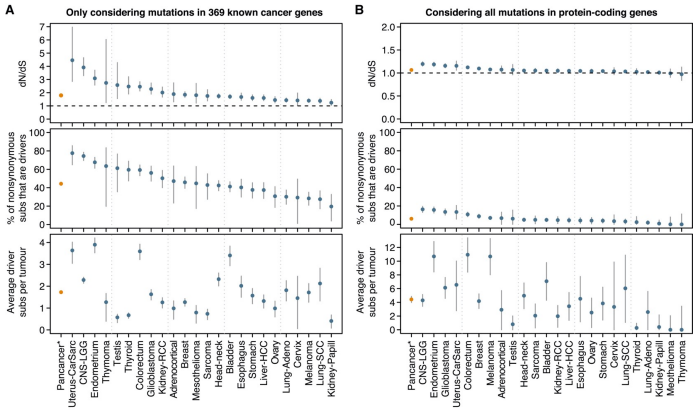


FIG. 1.—Change in mortality with age for cancer of the oesophagus, stomach and pancreas in men and for cancer of the stomach and pancreas in women shown on a double logarithmic scale, that is, the logarithm of the death rate per million persons plotted against the logarithm of the mid-point of the age group. The straight line through the points has been drawn arbitrarily to give the best fit, subject to the gradient being 6 to 1.

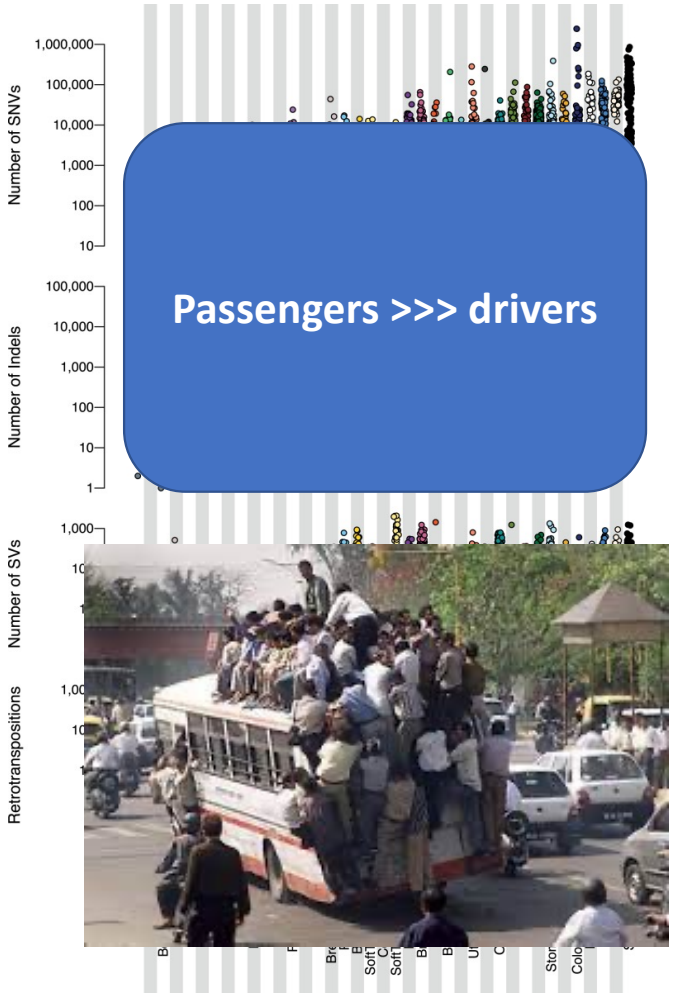
Armitage and Doll, 1954,
British Journal of Cancer



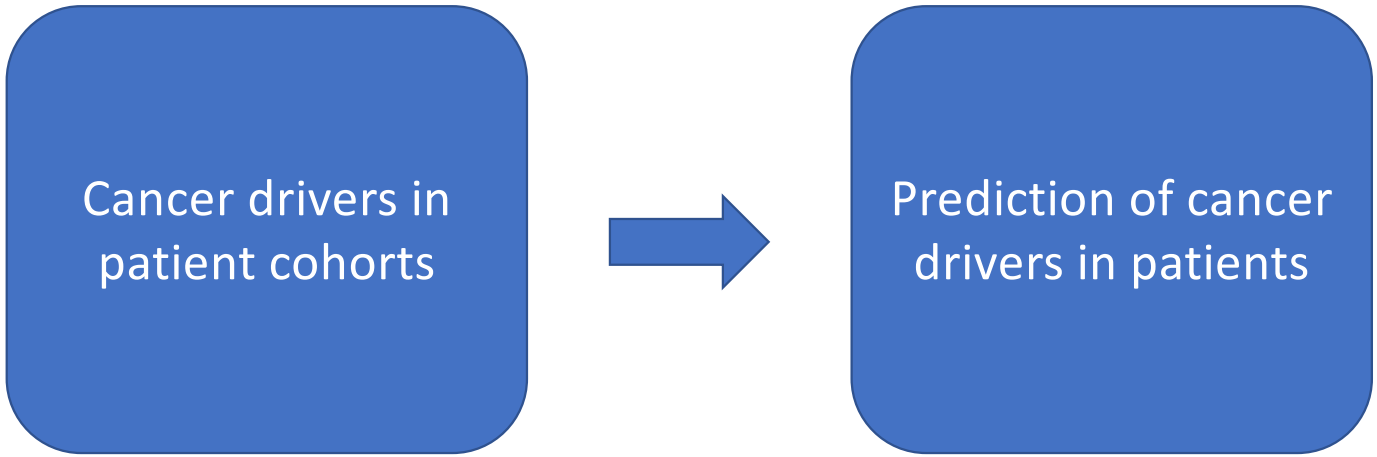
Martincorena et al, 2017, *Cell*

Typically, a handful of drivers per tumour

PCAWG, 2020, *Nature*



Research on cancer drivers



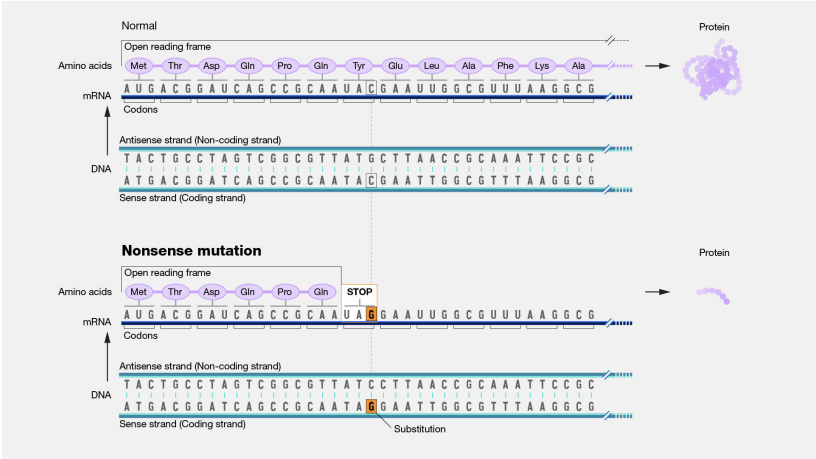
General Biology
Cancer Biology
New drivers
Cohort characterisation
Drug research

Risk prediction
Treatment

Types of point mutations: substitutions and indels

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	**UAA Stop**	**UGA Stop**	A
	UUG Leu	UCG Ser	**UAG Stop**	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

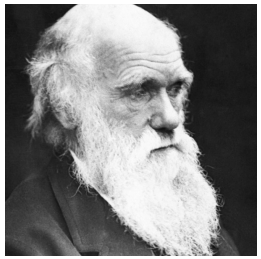
Nonsense substitution

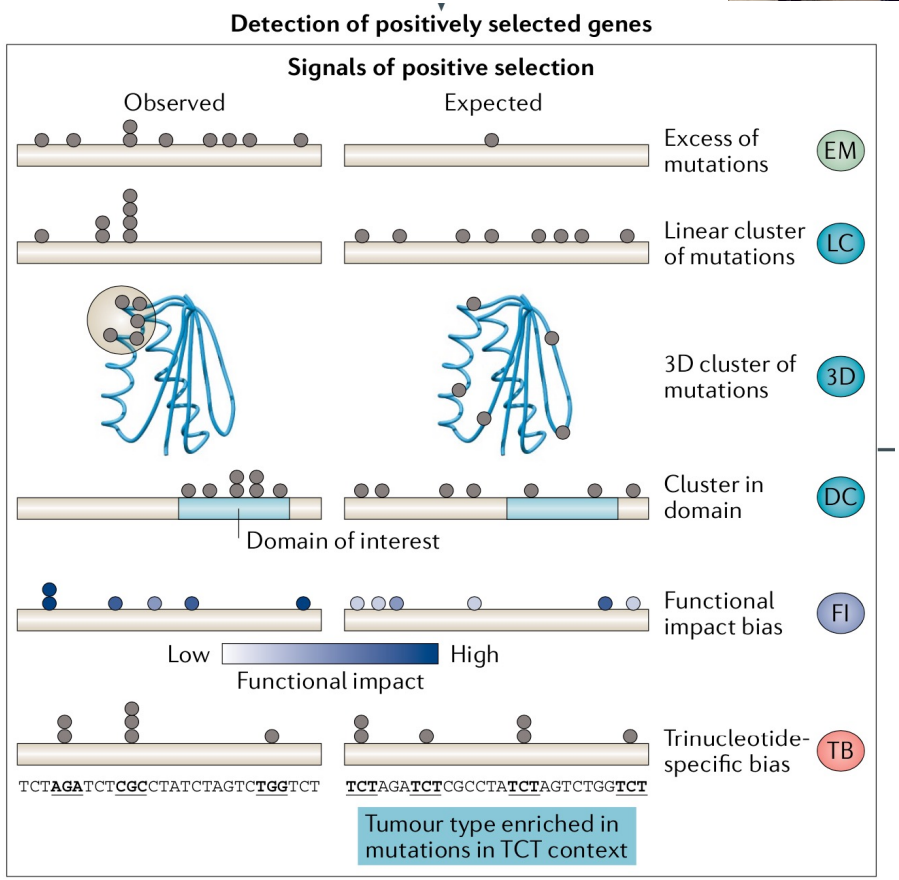


Synonymous substitution, e.g. TAT > TAC (Tyr → Tyr)
Missense substitution, e.g. TAT > TGT (Tyr → Cys)
Nonsense substitution, e.g. TAT > TAA (Tyr → Stop*)
Indels: insertions/deletions, in frame *vs* out of frame

How to find drivers in cohorts?

- Like finding needles in a haystack
- Recurrence – signature of positive selection
- Key – properly modelling the mutational process:
  null (”expected”) model



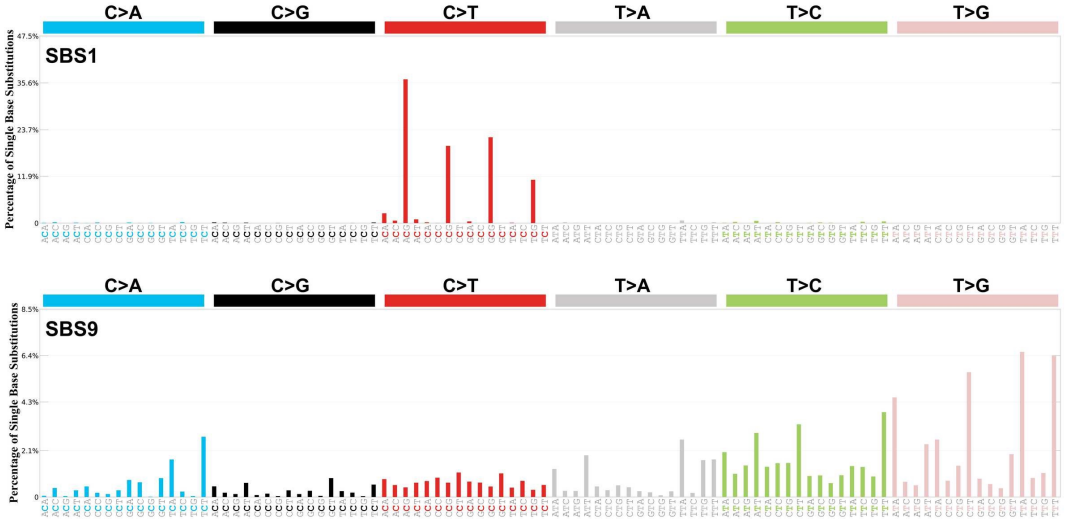


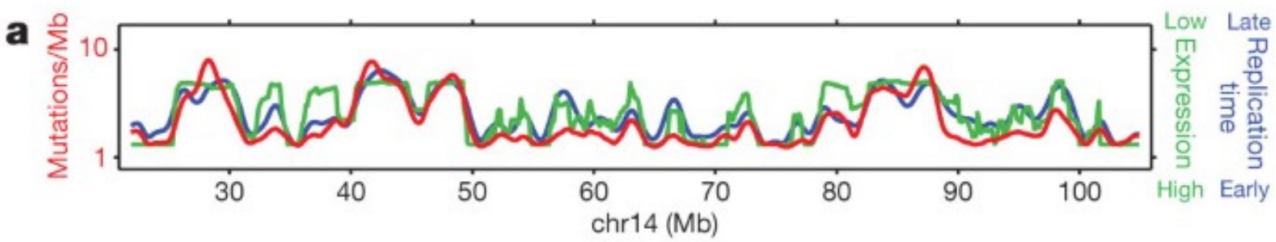
Martínez-Jiménez et al, 2020, *Nature Reviews Cancer*

Modelling the mutational process

- Null model
- Difficult for:
  - structural variants
    - Gistic https://www.genepattern.org/modules/docs/GISTIC_2.0
    - Gene fusions, etc
  - epigenetic alterations
- Better understood for point mutations
  - substitutions
  - small indels

Most substitution mutational processes can be described using the trinucleotide context





Lawrence et al, 2013, *Nature*

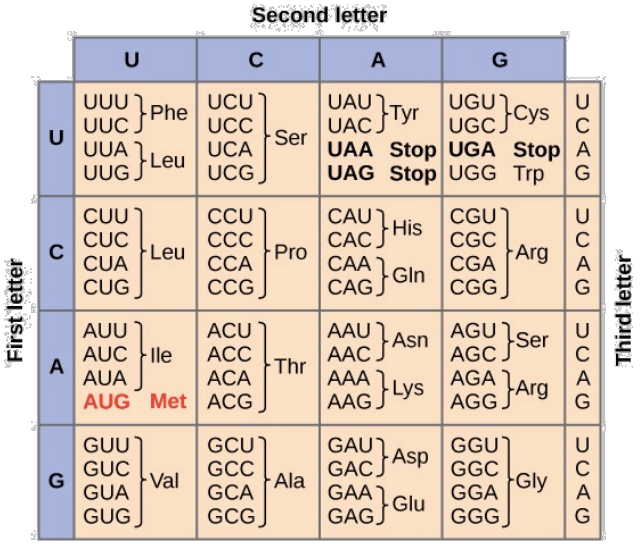
Selection for substitutions in coding regions

- dN/dS: a formal observed/expected test
  - Rate of non-synonymous substitutions divided by the rate of synonymous substitutions (=obs /exp under null model)
  - dN/dS < 1? **Negative selection**
  - dN/dS = 1? **Neutral evolution**
  - dN/dS > 1? **Positive selection!**
  - `dndscv` https://github.com/im3sanger/dndscv

- Non-synonymous substitutions:
  - Missense (e.g. Leu → Pro)
  - Nonsense (e.g. Ser → Stop)
  - Splice sites

(OGs and TSGs)

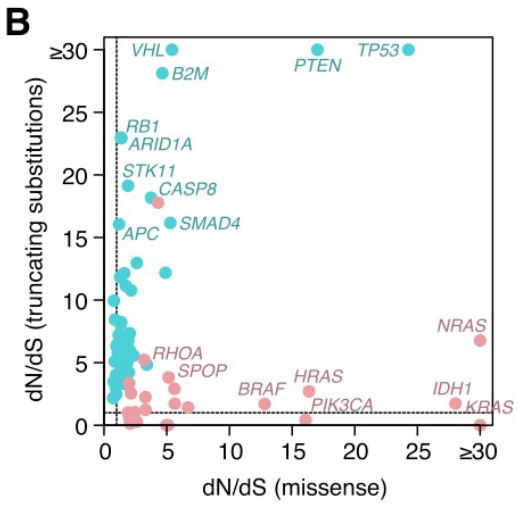
Truncating (TSGs)



Selection for point mutations in coding regions

- Oncogenes vs tumour suppressors:
  - OG: missense
  - TSG: nonsense, splice, indels, missense

- Coefficient of selection (dN/dS) interpretation
  - dN/dS of 1: all obs non-syn are exp (neutral)
  - dN/dS of 2: 50% of non-syn selected
  - dN/dS of 10: 90% of non-syn selected
  - dN/dS of 100: 99% of non-syn selected



Martincorena et al, 2017, *Cell*

dndscv detecting selection

- R package dndscv (Martincorena et al, *Cell* 2017)
  - dN/dS model – coef. of selection
  - Sophisticate mutational model:
    - Trinuc frequencies
    - Covariates for regional variation + data at hand
    - Overdispersion – negative binomial (equivalent of Poisson distribution)
  - User friendly
  - https://github.com/im3sanger/dndscv
- Selection estimates at:
  - Gene level (or domains)
  - Global
  - Sites & codons
  - Not only for cancer

`dndscv` gene level

- How to run:
  - `dndscv` function
  - panel vs no panel
  - Hypermutators
- Quick tutorial:
  http://htmlpreview.github.io/?http://github.com/im3sanger/dndscv/blob/master/vignettes/dNdScv.html
- Outputs:

```
dout = dndscv(muts)
names(dout)
 [1] "globaldnds"   "sel_cv"       "sel_loc"      "annotmuts"    "genemuts"
 [6] "geneindels"   "mle_submodel" "exclsamples"  "exclmuts"     "nbreg"
[11] "nbregind"     "poissmodel"   "wrongmuts"    "N"            "L"
```

`dndscv$sel_cv` example: TCGA bladder carcinoma

gene_name	n_syn	n_mis	n_non	n_spl	n_ind	wmis_cv	wnon_cv	wspl_cv	wind_cv	qglobal_cv
TP53	5	143	35	4	24	41.0	97.7	97.7	366.9	0
PIK3CA	3	80	0	0	1	14.5	0.0	0.0	6.1	0
FGFR3	6	53	1	0	2	12.6	2.6	2.6	13.7	0
ARID1A	7	34	45	6	30	2.7	38.1	38.1	65.7	0
KDM6A	6	21	33	9	34	2.5	33.6	33.6	136.9	0
KMT2D	9	39	50	10	29	1.7	23.6	23.6	26.2	0
RB1	0	8	34	14	21	2.3	91.4	91.4	138.5	0
STAG2	8	12	22	4	14	1.3	21.8	21.8	64.8	0
ELF3	0	29	1	1	20	17.1	10.1	10.1	138.9	0
CDKN2A.p16	0	12	3	2	5	16.2	104.4	104.4	188.4	0
CDKN1A	1	5	6	0	23	5.6	51.2	51.2	389.1	0
TSC1	3	9	12	5	6	1.4	25.6	25.6	27.1	1.86E-13
ZFP36L1	0	7	3	0	18	4.1	29.5	29.5	85.4	1.54E-12
FBXW7	0	19	10	1	2	7.5	35.3	35.3	16.8	5.32E-10
EP300	2	40	16	1	7	4.5	14.4	14.4	14.1	2.90E-09

w = dN/dS (coef. of selection)
(w-1)/w = fraction mutations selected

TP53 wmis = 41.0, 97.5% selected
EP300 wmis = 4.5, 77.8% selected

Why are coefficients of selection important?

- Personalised medicine implications
- Hypermutators lower dN/dS

…

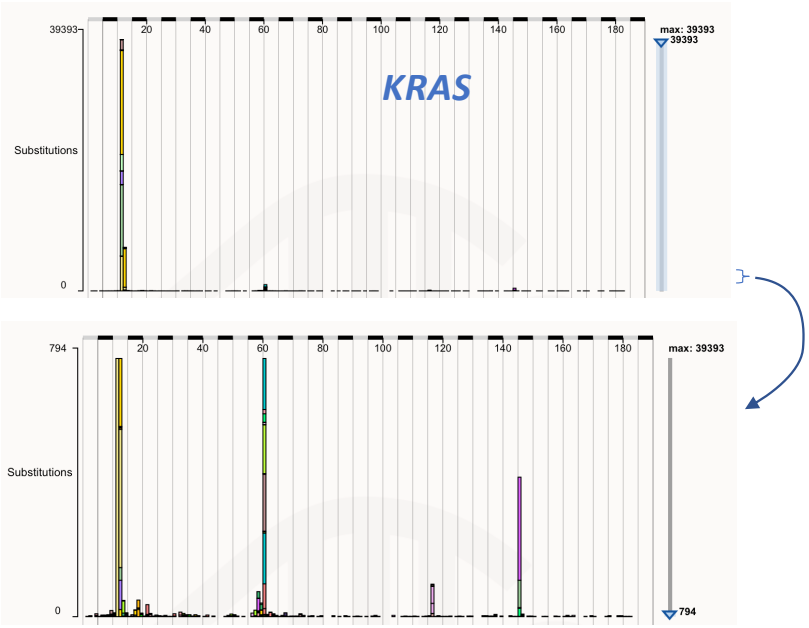
`dndscv` tutorials:
https://drive.google.com/drive/folders/1-M4eelsdzqZCyEwEnFNHFDZs2LkZAOdr?usp=sharing
https://github.com/im3sanger/dndscv

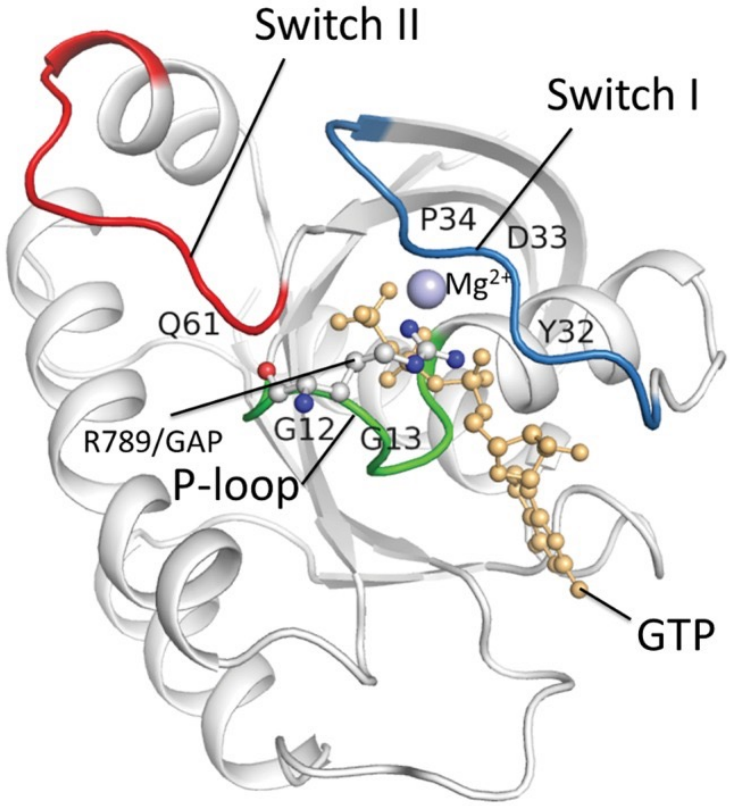
Other tools:

- `MutSigCV`: https://www.genepattern.org/modules/docs/MutSigCV
- `IntOGen`: https://www.intogen.org/search

dndscv detecting selection at sites/codons

- Recurrence - hotspots
- COSMIC:





Chen et al, 2013, *PLoS one*

Hotspot analysis with `dndscv`: `sitednds` and `codondnds`

Site dN/dS on TCGA data

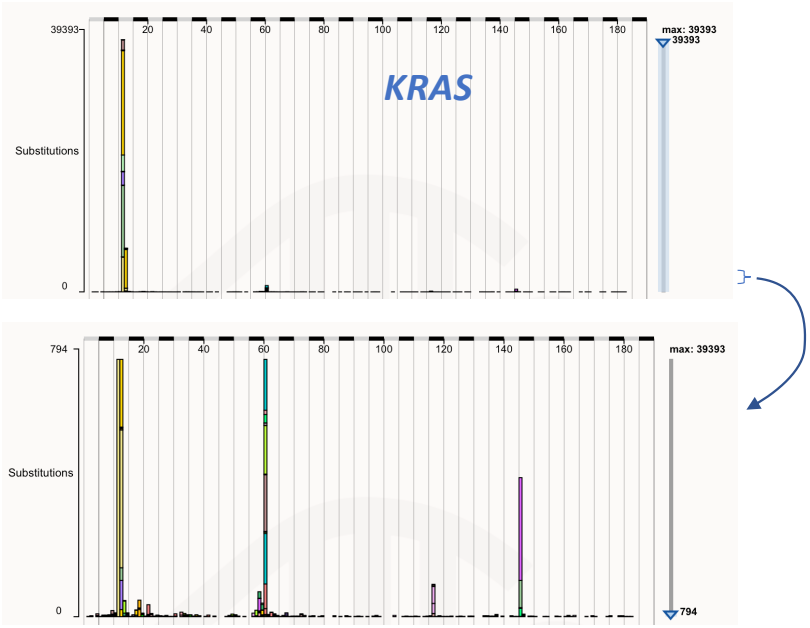
chr	pos	ref	mut	gene	aachange	impact	ref3_cod	mut3_cod	freq	mu	dnds	pval	qval
7	140453136	A	T	BRAF	V600E	Missense	GTG	GAG	408	0.002	176526.1	0	0
2	209113112	C	T	IDH1	R132H	Missense	CGT	CAT	353	0.079	4469.6	0	0
3	178952085	A	G	PIK3CA	H1047R	Missense	CAT	CGT	161	0.007	22252.8	0	0
12	25398284	C	T	KRAS	G12D	Missense	GGT	GAT	114	0.008	14578.7	6.21E-241	1.62E-233
3	178936091	G	A	PIK3CA	E545K	Missense	TGA	TAA	150	0.029	5112.8	4.94E-232	1.03E-224
12	25398284	C	A	KRAS	G12V	Missense	GGT	GTT	95	0.007	13634.6	1.26E-205	2.19E-198
1	115256529	T	C	NRAS	Q61R	Missense	CAA	CGA	72	0.003	25980.9	1.02E-184	1.51E-177
3	178936082	G	A	PIK3CA	E542K	Missense	TGA	TAA	95	0.029	3238.1	3.17E-147	4.13E-140
17	7578406	C	T	TP53	R175H	Missense	CGC	CAC	108	0.056	1938.8	2.16E-138	2.50E-131
12	25398285	C	A	KRAS	G12C	Missense	TGG	TTG	62	0.007	8299.3	1.36E-132	1.42E-125
17	7577538	C	T	TP53	R248Q	Missense	CGG	CAG	92	0.045	2025.9	9.10E-126	8.62E-119
17	7577121	G	A	TP53	R273C	Missense	GCG	GTG	90	0.056	1596.8	5.13E-115	4.46E-108
1	115256530	G	T	NRAS	Q61K	Missense	ACA	AAA	48	0.004	11185.0	2.62E-114	2.10E-107
17	7578190	T	C	TP53	Y220C	Missense	TAT	TGT	49	0.015	3290.9	2.00E-90	1.49E-83
4	1803568	C	G	FGFR3	S249C	Missense	TCC	TGC	35	0.004	8415.1	5.05E-84	3.51E-77
17	7577094	G	A	TP53	R282W	Missense	CCG	CTG	60	0.049	1216.8	3.12E-80	1.91E-73
17	7577539	G	A	TP53	R248W	Missense	CCG	CTG	60	0.049	1216.8	3.12E-80	1.91E-73
11	533874	T	C	HRAS	Q61R	Missense	CAG	CGG	29	0.002	14650.1	4.55E-79	2.63E-72
17	7577120	C	T	TP53	R273H	Missense	CGT	CAT	69	0.083	832.5	1.53E-77	8.36E-71
12	25398285	C	G	KRAS	G12R	Missense	TGG	TCG	28	0.002	14496.8	1.15E-76	5.99E-70
12	25398281	C	T	KRAS	G13D	Missense	GGC	GAC	34	0.008	4430.7	1.19E-72	5.90E-66
17	7578394	T	C	TP53	H179R	Missense	CAT	CGT	32	0.008	3808.1	3.61E-67	1.71E-60
3	178952085	A	T	PIK3CA	H1047L	Missense	CAT	CTT	26	0.003	8447.0	5.41E-66	2.45E-59

…. 565 with q<0.1

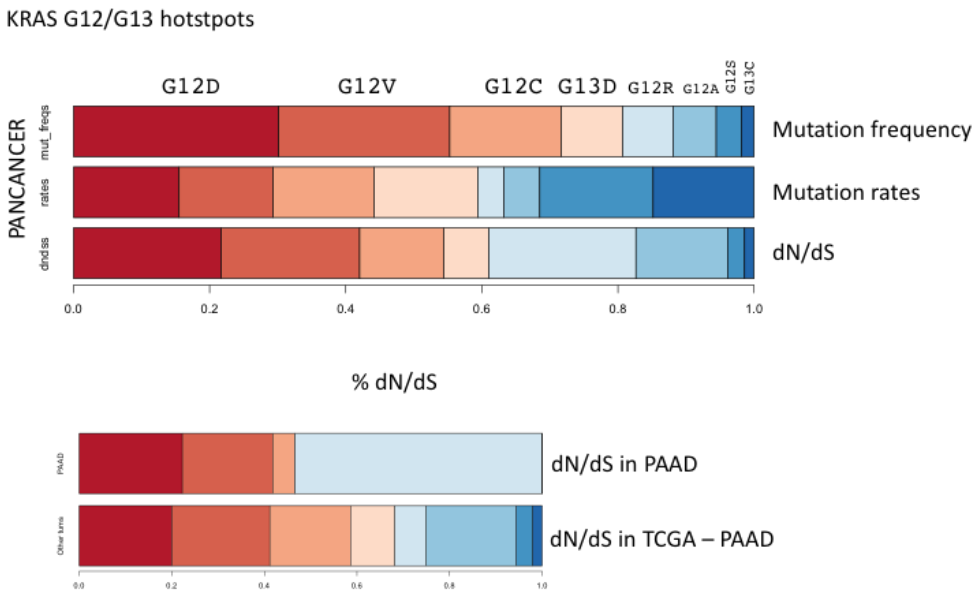
Quick tutorial: https://rdrr.io/github/im3sanger/dndscv/f/vignettes/sitednds.Rmd

Hotspots: `sitednds` and `codondnds`

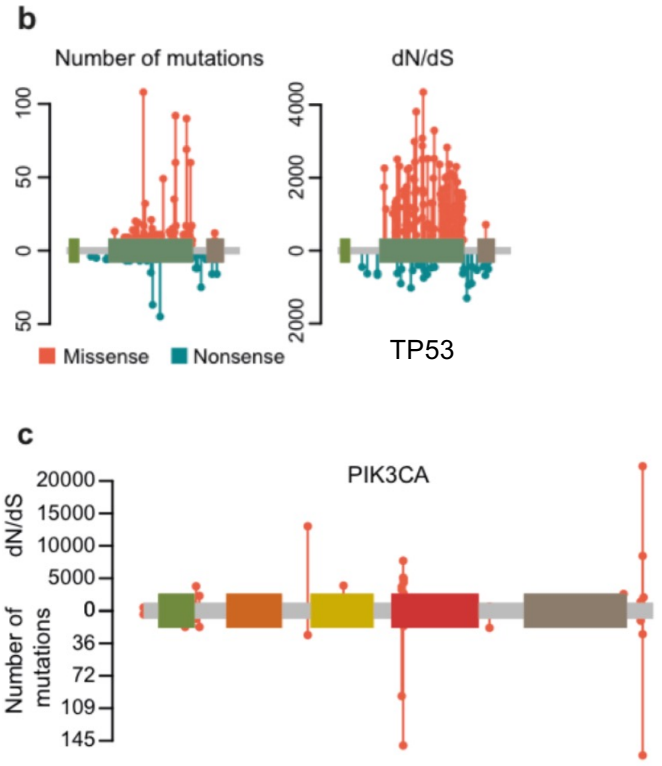
- Recurrence
- COSMIC:

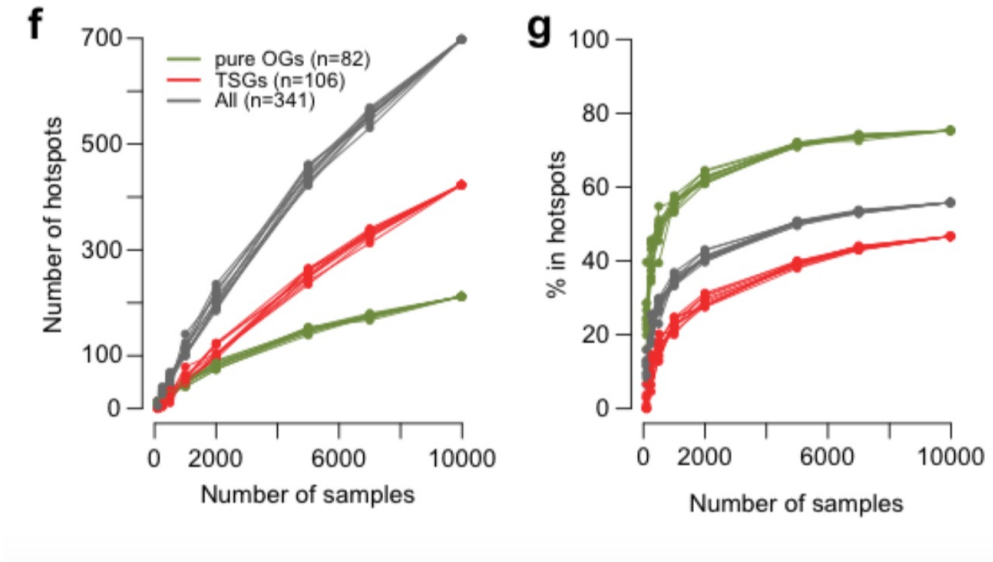


KRAS hotspots vary from tumour to tumour



Hotspots: best for OGs





dndscv example: drivers in healthy bladder

Andrew Lawson

RESEARCH

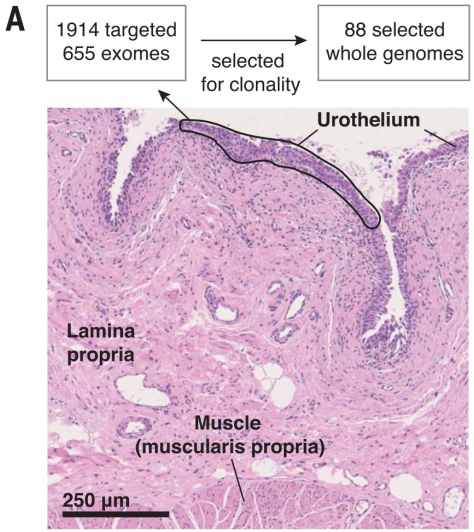
RESEARCH ARTICLE

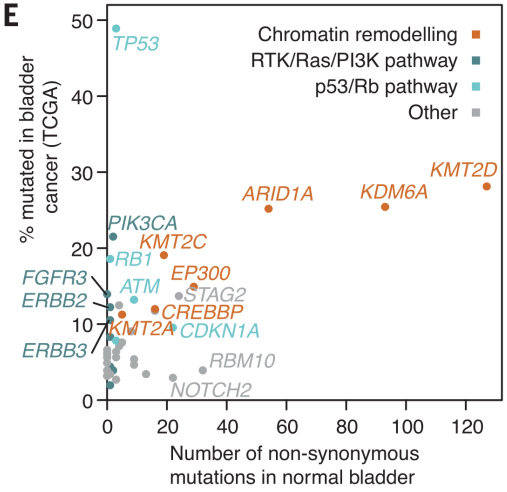
MUTATION

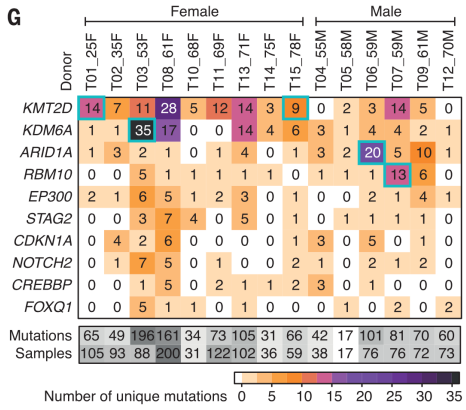
Extensive heterogeneity in somatic mutation and selection in the human bladder

Andrew R. J. Lawson[1], Federico Abascal[1], Tim H. H. Coorens[1], Yvette Hooks[1], Laura O'Neill[1], Calli Latimer[1], Keiran Raine[1], Mathijs A. Sanders[2,3], Anne Y. Warren[4], Krishnaa T. A. Mahbubani[4,5], Bethany Davidson[4,5], Timothy M. Butler[1], Luke M. R. Harvey[1], Alex Cagan[1], Andrew Menzies[1], Luiza Moore[1,3], Alexandra J. Colquhoun[6], William Turner[6], Benjamin Thomas[1,6], Vincent Gnanapragasam[7,8], Nicholas Williams[1], Doris M. Rassl[9], Harald Vöhringer[10], Sonia Zumalave[11], Jyoti Nangalia[1], José M. C. Tubio[12,13,11], Moritz Gerstung[10], Kourosh Saeb-Parsy[4,5], Michael R. Stratton[1], Peter J. Campbell[1,14], Thomas J. Mitchell[4,5], Iñigo Martincorena[1]

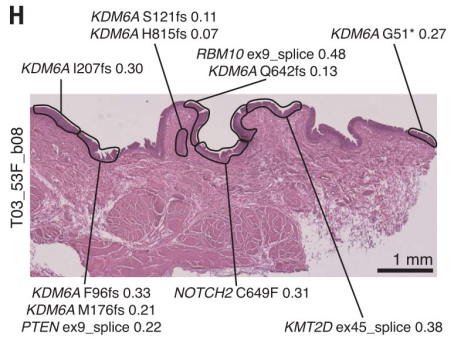
Lawson et al, 2020, *Science*

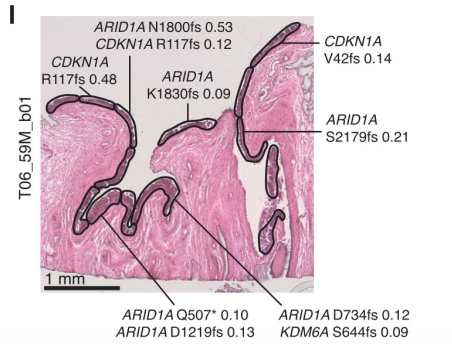






- *Chromatin modifiers!*
- *TP53 early diagnosis?*
- *Differences between donors*





Regulatory regions and noncoding genes

- Protein coding: 1% of the genome
- PCAWG: 2,658 whole genomes (TCGA exomes)
  - microRNAs, lncRNAs, tRNAs…
  - Promoters, 3’ and 5’ UTRs, enhancers

Article

Analyses of non-coding somatic drivers in 2,658 cancer whole genomes

https://doi.org/10.1038/s41586-020-1965-x
Received: 19 January 2018
Accepted: 2 December 2019
Published online: 5 February 2020
Open access

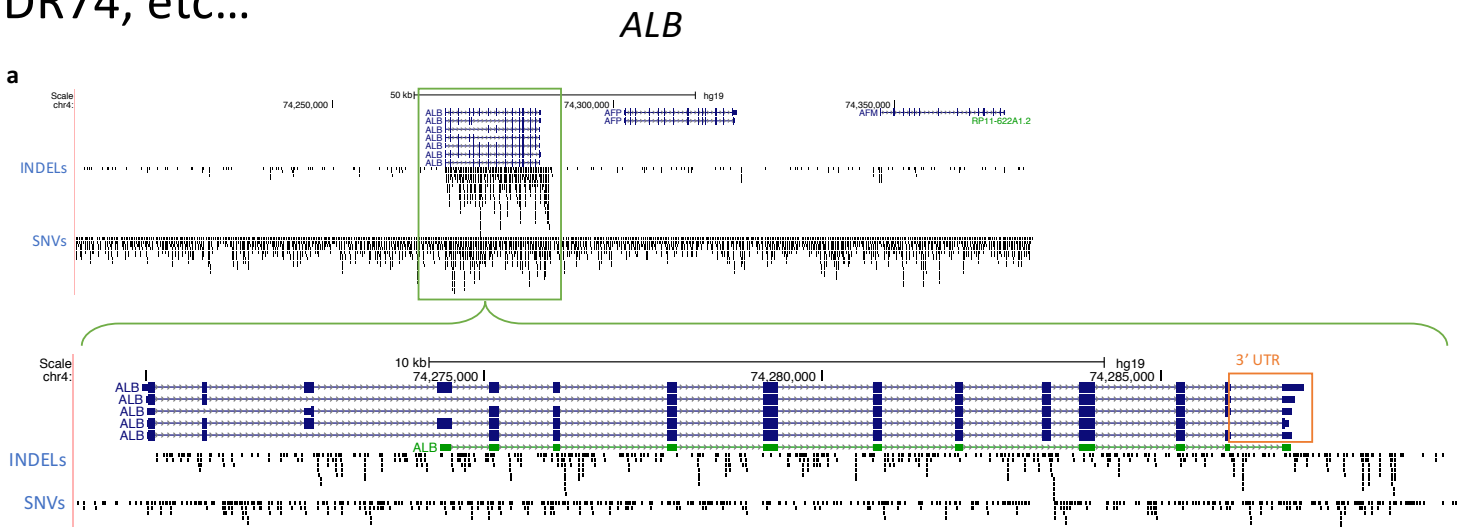
Esther Rheinbay[1,2,3], Morten Muhlig Nielsen[4,3], Federico Abascal[5,6], Jeremiah A. Wala[7,8,9], Ofer Shapira[2,3,10], Grace Tiao[1], Henrik Hornshøj[4], Julian M. Hess[1], Randi Istrup Juul[4], Ziao Lin[11], Lars Feuerbach[12], Radhakrishnan Sabarinathan[12,13], Tobias Madsen[4], Jaegil Kim[1], Loris Mularoni[12,13], Shimin Shuai[12,13], Andrés Lanzós[14,15,16], Carl Herrmann[17,18], Yosef E. Maruvka[19], Ciyue Shen[20,21], Samirkumar B. Amin[22,23], Pratiti Bandopadhayay[1,2], Johanna Bertl[4], Keith A. Boroevich[24], John Busanovich[1,2], Joana Carlevaro-Fita[25,26,27], Dimple Chakravarty[28,29], Calvin Wing Yiu Chan[30,31], David Craft[32], Priyanka Dhingra[28,29], Klev Diamanti[33], Nuno A. Fonseca[34], Abel Gonzalez-Perez[35,36], Qianyun Guo[37], Mark P. Hamilton[38], Nicholas J. Haradhvala[1,2], Chen Hong[8,39], Keren Isaev[40,41], Todd A. Johnson[23], Malene Juul[4], Andre Kahles[42], Abdullah Kahraman[42], Youngwook Kim[37], Jan Komorowski[43,44], Kiran Kumar[17], Sushant Kumar[45], Donghoon Lee[46], Kjong-Van Lehmann[47], Yilong Li[48,49], Eric Minwei Liu[28,29], Lucas Lochovsky[42], Keunchil Park[50], Oriol Pich[51,52], Nicola D. Roberts[17], Gordon Saksena[1], Steven E. Schumacher[1,2], Nikos Sidiropoulos[4,53], Lina Sieverling[54,55], Nasa Sinnott-Armstrong[56,57], Chip Stewart[1], David Tamborero[58,59], Jose M. C. Tubio[14,60,61], Husen M. Umer[62], Lis Uusküla-Reimand[48,49], Claes Wadelius[62], Lina Wadi[17], Xiaotong Yao[63], Cheng-Zhong Zhang[19,64], Jing Zhang[65], James E. Haber[66], Asger Hobolth[33], Marcin Imielinski[67,68], Manolis Kellis[1,69], Michael S. Lawrence[1,2], Christian von Mering[70], Hidewaki Nakagawa[71], Benjamin J. Raphael[72], Mark A. Rubin[14,61,62], Chris Sander[28,73], Lincoln D. Stein[12,74], Joshua M. Stuart[37], Tatsuhiko Tsunoda[75,76,77], David A. Wheeler[65], Rory Johnson[78,79], Jüri Reimand[12,80], Mark Gerstein[81,82,83], Ekta Khurana[28,29,84,85], Peter J. Campbell[3,86], Núria López-Bigas[58,59,87], PCAWG Drivers and Functional Interpretation Working Group[88], PCAWG Structural Variation Working Group[89], Joachim Weischenfeldt[4,5,90,91], Rameen Beroukhim[1,8,92,93], Iñigo Martincorena[3,94], Jakob Skou Pedersen[4,53,95], Gad Getz[1,2,3,27,96] & PCAWG Consortium[97]

Rheinbay et al, 2020, *Nature*

- Large set of driver candidates
  - *ALB*
  - lncRNAs: *NEAT1* & *MALAT1*…
  - UTRs/promoters: *WDR74*…
  - tRNAs
  - small RNAs
  - micro RNAs

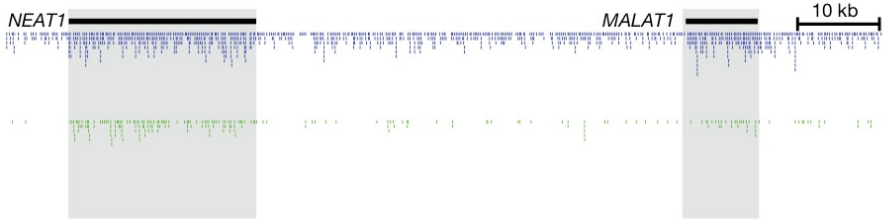
Regulatory and noncoding genes

- Preliminary set of candidates
  - ~~*ALB*~~
  - lncRNAs: NEAT1 & MALAT1…
  - UTRs/promoters: WDR74, etc…
  - tRNAs
  - small RNAs
  - micro RNAs

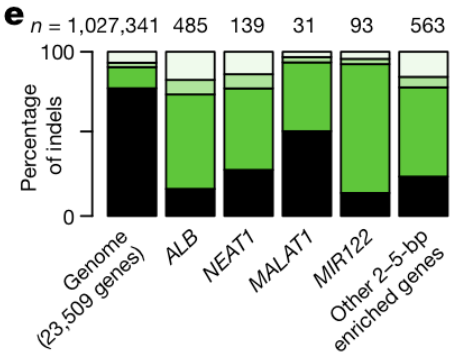


Regulatory and noncoding genes

- Preliminary set of candidates
  - ~~*ALB*~~
  - ~~lncRNAs: NEAT1 & MALAT1…~~
  - UTRs/promoters: WDR74, etc…
  - tRNAs
  - small RNAs
  - micro RNAs

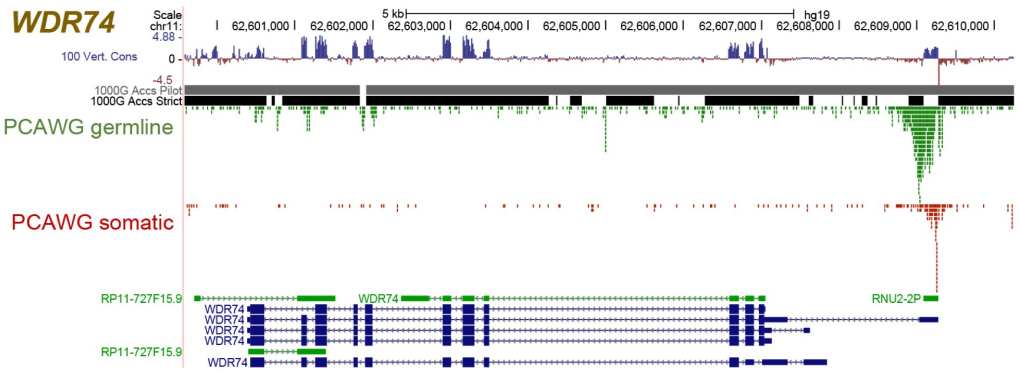


New unaccounted mutational processes
(null model violation)



Regulatory and noncoding genes

- Preliminary set of candidates
  - ~~*ALB*~~
  - ~~lncRNAs: NEAT1 & MALAT1…~~
  - ~~UTRs/promoters: WDR74, etc…~~
  - ~~tRNAs~~
  - ~~small RNAs~~
  - ~~micro RNAs~~



Mapping problems

WDR74 had been reported in several publications

Regulatory regions of cancer genes:
promoters and UTRs

Article

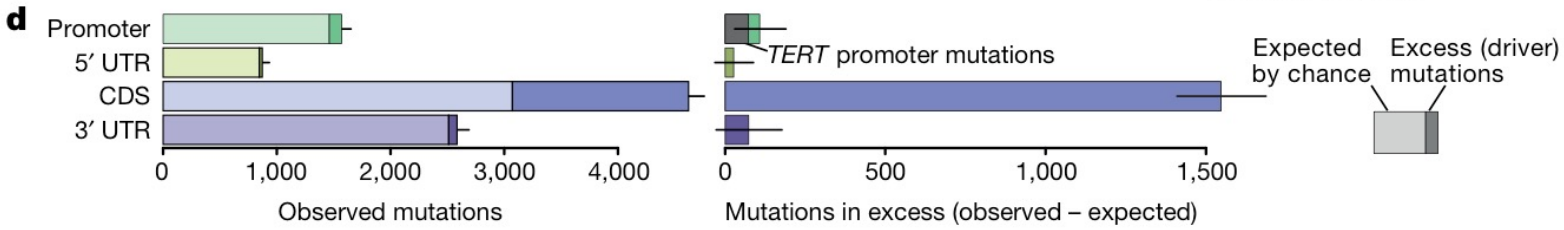
Analyses of non-coding somatic drivers in 2,658 cancer whole genomes

https://doi.org/10.1038/s41586-020-1965-x
Received: 19 January 2018
Accepted: 2 December 2019
Published online: 5 February 2020
Open access

Esther Rheinbay[1,2,3,77], Morten Muhlig Nielsen[4,5], Federico Abascal[6,7], Jeremiah A. Wala[1,8,77], Ofer Shapira[1,2,77], Grace Tiao[1], Henrik Hornshøj[4], Julian M. Hess[1], Randi Istrup Juul[4], Ziao Lin[1,9], Lars Feuerbach[10], Radhakrishnan Sabarinathan[10,11], Tobias Madsen[4], Jaegil Kim[1], Loris Mularoni[10,12], Shimin Shuai[10,13], Andrés Lanzós[10,14,15], Carl Herrmann[10,16], Yosef E. Maruvka[1,2], Ciyue Shen[10,20], Samirkumar B. Amin[17,18], Pratiti Bandopadhayay[19], Johanna Bertl[4], Keith A. Boroevich[20], John Busanovich[21], Joana Carlevaro-Fita[10,22,23], Dimple Chakravarty[24,25], Calvin Wing Yiu Chan[17,26], David Craft[21], Priyanka Dhingra[10,20], Klev Diamanti[20], Nuno A. Fonseca[27], Abel Gonzalez-Perez[10,11], Qianyun Guo[28], Mark P. Hamilton[29], Nicholas J. Haradhvala[1,2], Chen Hong[4,30], Keren Isaev[31,24], Todd A. Johnson[20], Malene Juul[4], Andre Kahles[32], Abdullah Kahraman[33], Youngwook Kim[34], Jan Komorowski[35,36], Kiran Kumar[1,37], Sushant Kumar[38], Donghoon Lee[39], Kjong-Van Lehmann[32], Yilong Li[40,41], Eric Minwei Liu[20,26], Lucas Lochovsky[42], Keunchil Park[43], Oriol Pich[10,11], Nicola D. Roberts[44], Gordon Saksena[1], Steven E. Schumacher[1,2], Nikos Sidiropoulos[45], Lina Sieverling[3,46], Nasa Sinnott-Armstrong[1,47], Chip Stewart[1], David Tamborero[10,11], Jose M. C. Tubio[48,49,50], Husen M. Umer[20], Liis Uusküla-Reimand[48,49], Claes Wadelius[30], Lina Wadi[51], Xiaotong Yao[52], Cheng-Zhong Zhang[53,54], Jing Zhang[38], James E. Haber[34], Asger Hobolth[55], Marcin Imielinski[27,56], Manolis Kellis[1,45], Michael S. Lawrence[57], Christian von Mering[58], Hidewaki Nakagawa[59], Benjamin J. Raphael[60], Mark A. Rubin[10,61,62], Chris Sander[19,32], Lincoln D. Stein[63,64], Joshua M. Stuart[65], Tatsuhiko Tsunoda[66,67,68], David A. Wheeler[69], Rory Johnson[70,71], Jüri Reimand[72,73], Mark Gerstein[74,42,68], Ekta Khurana[20,75,76,77], Peter J. Campbell[1,2,8], Núria López-Bigas[10,11,12], PCAWG Drivers and Functional Interpretation Working Group[78], PCAWG Structural Variation Working Group[79], Joachim Weischenfeldt[42,68,74a], Rameen Beroukhim[1,8,78,79a], Iñigo Martincorena[6,79a], Jakob Skou Pedersen[4,30,79a], Gad Getz[1,2,3,77,79a] & PCAWG Consortium[72]

- No significant hits in cancer genes
  (except for *TERT* promoter, *TP53* 5'UTR and *TOB1* 3'UTR)

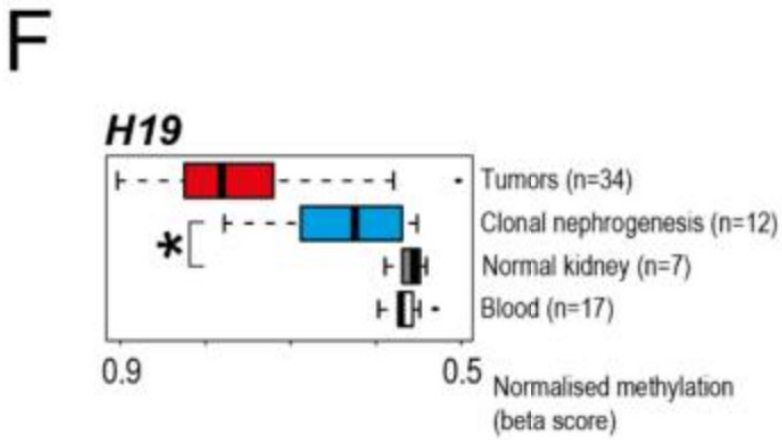
- Aggregated analysis (603 cancer genes)



- Not much beyond *TERT* promoter

Epigenetic drivers

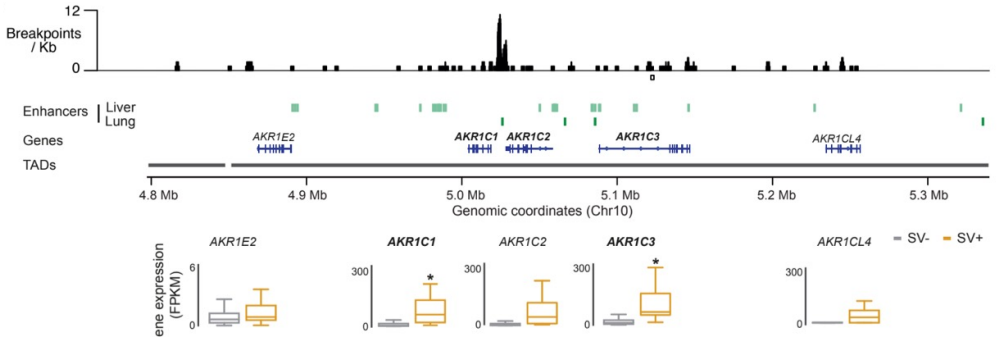
- Chromatin modifiers
- Plasticity: Δ individuals, cell types, environment → null model?
- Wilms tumour: *H19* hypermethylation. Noncoding driver!
- *VHL* silencing through hypermethylation



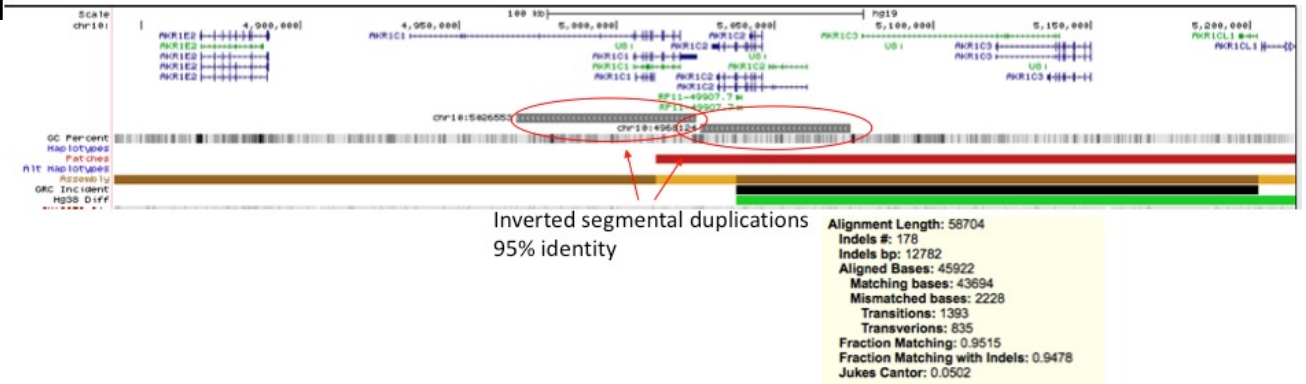
Coorens et al, 2019, *Science*

Structural drivers

- Copy number: GISTIC (Mermel et al, 2011): identifies regions of the genome that are significantly amplified or deleted across a set of samples (e.g. *MYC* oncogene amplification)
- Rearrangements:
  - Null model? Fragile sites *why?*
  - Gene fusions, enhancer hitch-hiking, copy number gai
    etc
  - AKR1C locus

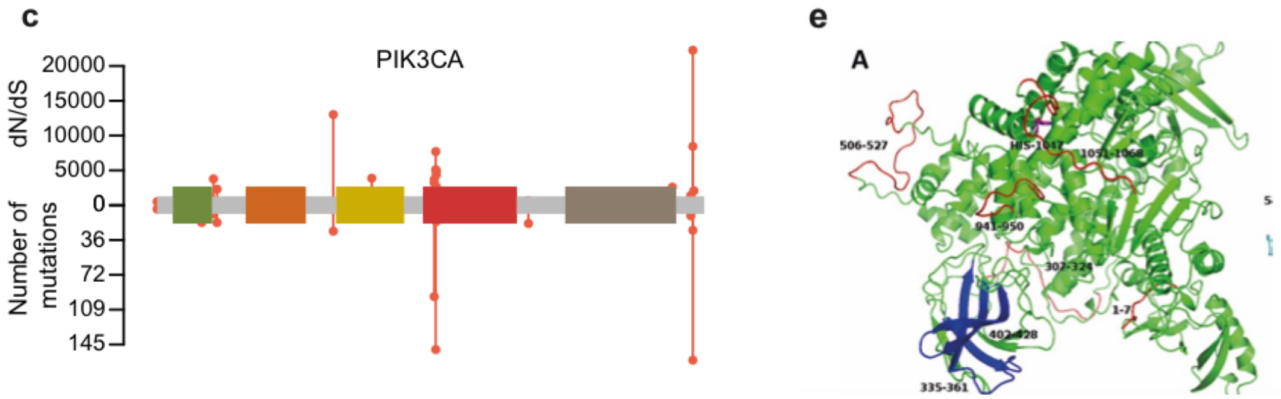


Rheinbay et al, 2020, *Nature*



Take home messages

- OG and TSG behave very differently – no B&W in Biology
- Passengers >>> drivers (hypermutators particularly problematic)
- Recurrence = positive selection (obs>exp)
- Most drivers are protein-coding (+TERT) – 1% of the genome!
  - Always look into the details
- Not all non-synonymous mutations are drivers (dN/dS)
- Structural and epigenetic alterations can be drivers too



Start codon loss in OG?

Prediction of driver mutations in a given patient

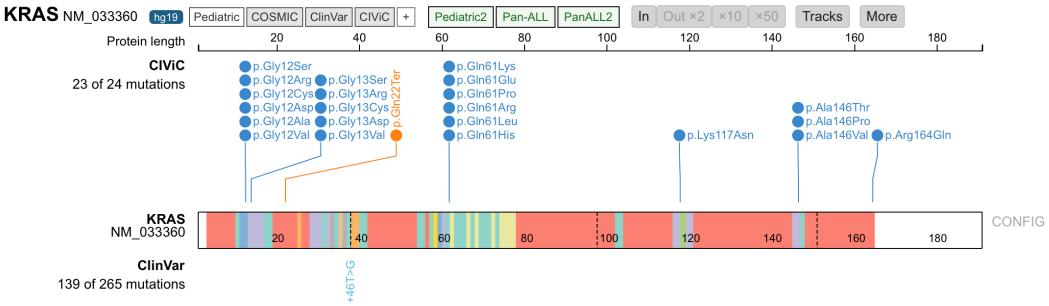
Table 1. Molecular targets for personalised cancer therapies

Cancer type	Cellular target	Targeted agent	Class of agent
Colorectal[16–18]	KRAS	Cetuximab	Monoclonal antibody against EGFR
Breast[19,20]	HER2	Trastuzumab	Monoclonal antibody against HER2/Neu (EGFR2)
Chronic myeloid leukaemia[21,22]	BCR-ABL fusion protein	Imatinib	Receptor tyrosine kinase inhibitor
Gastrointestinal stromal tumours[23,24]	c-KIT	Imatinib	Receptor tyrosine kinase inhibitor
Non-small-cell lung cancer[25–28]	EGFR	Erlotinib and gefitinib	Receptor tyrosine kinase inhibitor
Non-small-cell lung cancer[29,30]	EML4-ALK fusion protein	Crizotinib	Receptor tyrosine kinase inhibitor
Metastatic malignant melanoma[31,32]	BRAF V600E	Vemurafenib	B-raf/MEK/ERK pathway inhibitor
Ovarian, breast and prostate cancer (under investigation)[33,34]	BRCA1, BRCA2	Olaparib	Poly(ADP-ribose) polymerase (PARP) inhibitor

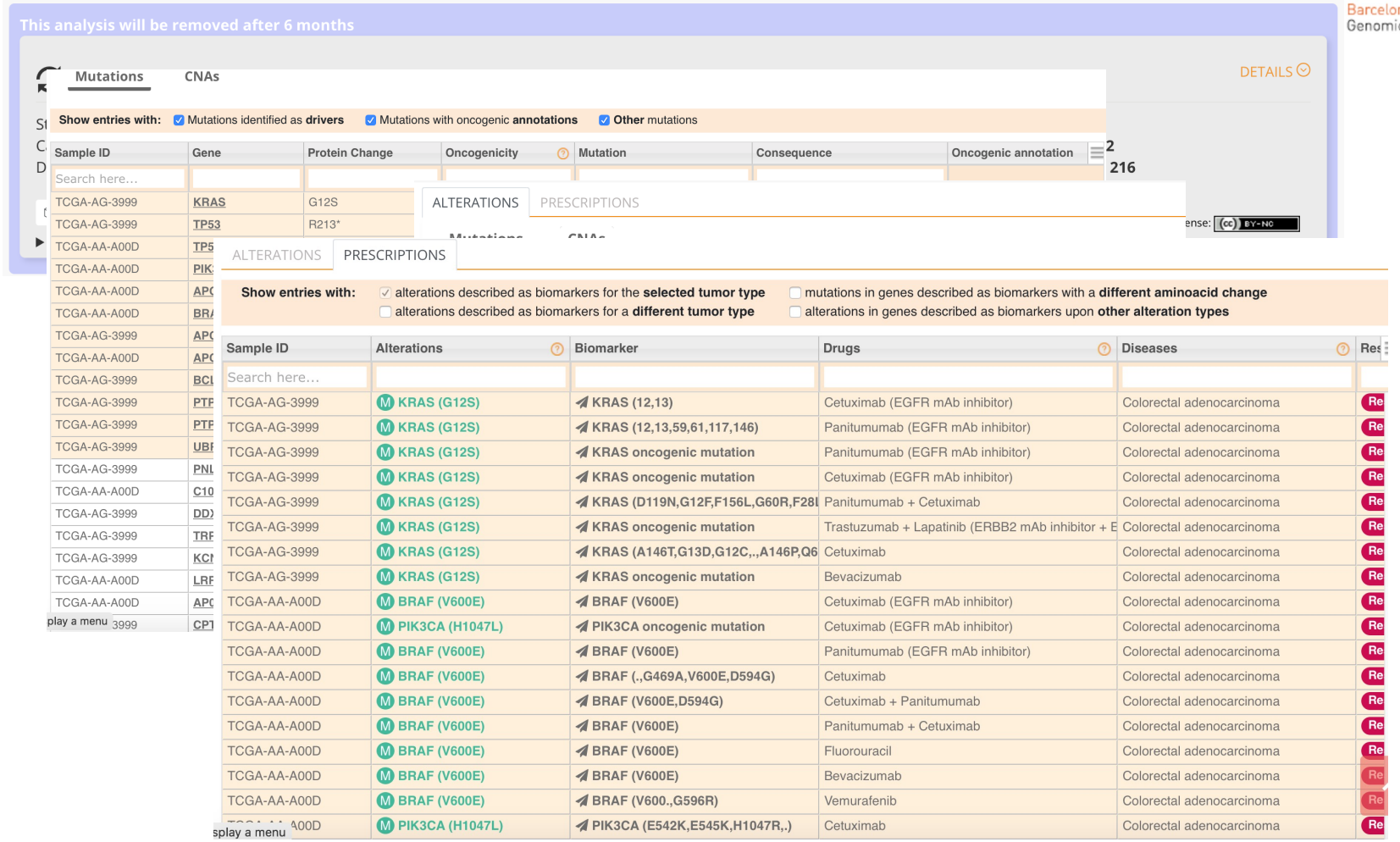
Abbreviations: APC: adenomatous polyposis coli; CML: chronic myeloid leukaemia; CRC: colorectal cancer; EGFR: epidermal growth factor receptor; EML4-ALK: echinoderm microtubule-associated protein-like 4—anaplastic lymphoma kinase fusion gene; FAP: familial adenomatous polyposis coli; GIST: gastrointestinal stromal tumour; NICE: National Institute for Health and Care Excellence; NSCLC: non-small cell lung cancer; PARP: poly(ADP-ribose) polymerase; TK: tyrosine kinase; TKI: tyrosine kinase inhibitor

Clinical practice

- Top genes: most mutations are drivers (*TP53*, *KRAS*, *BRAF*…)
  - Careful with hypermutators
- Hotspots: most drivers
- OGs and TSGs behave differently
- Useful tools:
  - CiVIC: https://civicdb.org/home
  - Cancer Genome Interpreter: https://www.cancergenomeinterpreter.org/home



Cancer Genome Interpreter example



BBGLab
Barcelona Biomedical Genomics Lab

Tamborero et al, 2018, *Genome Medicine*
Muiños et al, 2021, *Nature*
BoostDM: CiVIC, recurrence, conservation, …

Recommended readings and software

- Martincorena, Iñigo, et al. "Universal patterns of selection in cancer and somatic tissues." Cell 171.5 (2017): 1029-1041.
- Rheinbay, Esther, et al. "Analyses of non-coding somatic drivers in 2,658 cancer whole genomes." Nature 578.7793 (2020): 102-111.
- Lawson, Andrew RJ, et al. "Extensive heterogeneity in somatic mutation and selection in the human bladder." Science 370.6512 (2020): 75-82
- Tamborero, David, et al. "Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations." Genome medicine 10.1 (2018): 1-8
- Gonzalez-Perez, Abel, et al. "IntOGen-mutations identifies cancer drivers across tumor types." Nature methods 10.11 (2013): 1081-1082.
- Software: dndscv, maftools, IntoGen, Cancer Genome Interpreter

Thank you!