

Genome-Phenome analysis of neoplasia-related traits across the primate phylogeny

27-09-2021

Alejandro Valenzuela, David Juan, Gerard Muntané, Arcadi Navarro

Comparative and evolutionary genomics lab

Neoplasm proportions as phenotypes of interest

Started working with 3 main clusters of neoplasm rates:

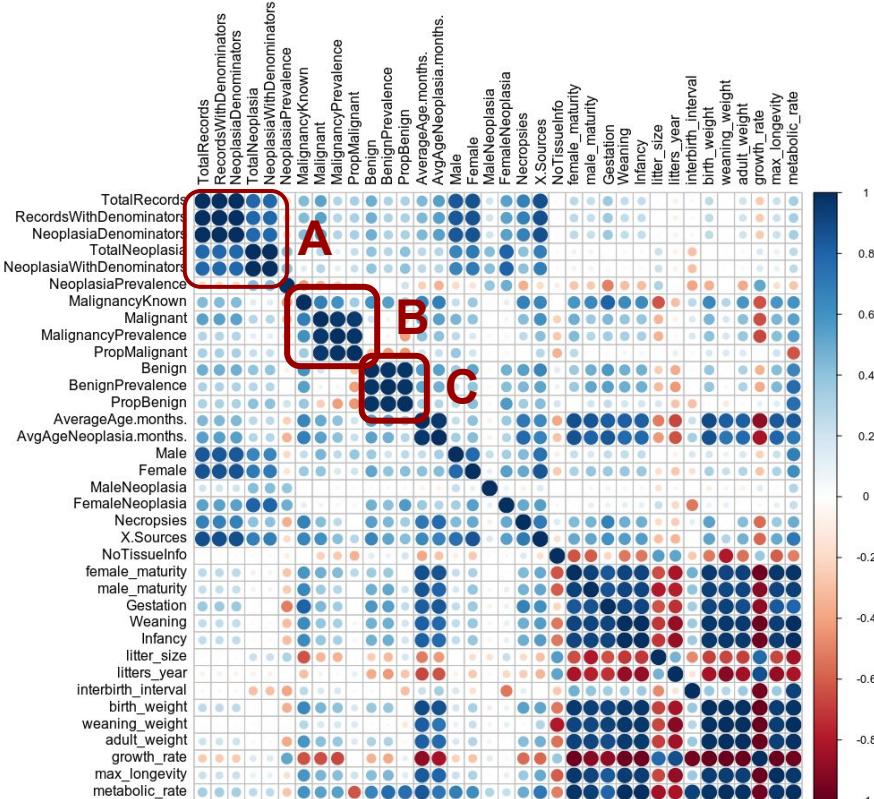
- A) General Neoplasia
- B) Malignancy
- C) Benignity

We focus on **protein-coding regions** and look at -only- two* potential genomic sources of phenotypic variation:

- Candidate Convergent Amino Acid Substitutions (CAAS).
- Correlations between rates of protein evolution and traits, using **Phylogenetic-Generalized Least Squares (PGLS)**

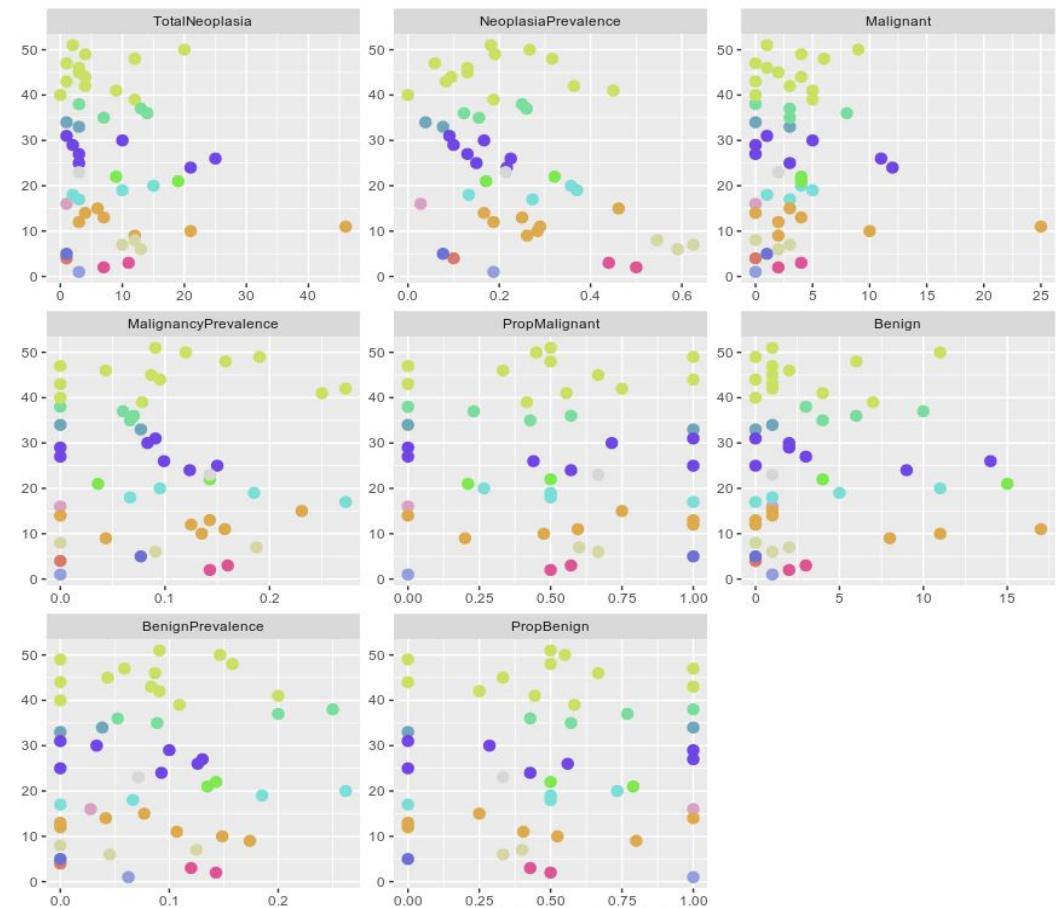
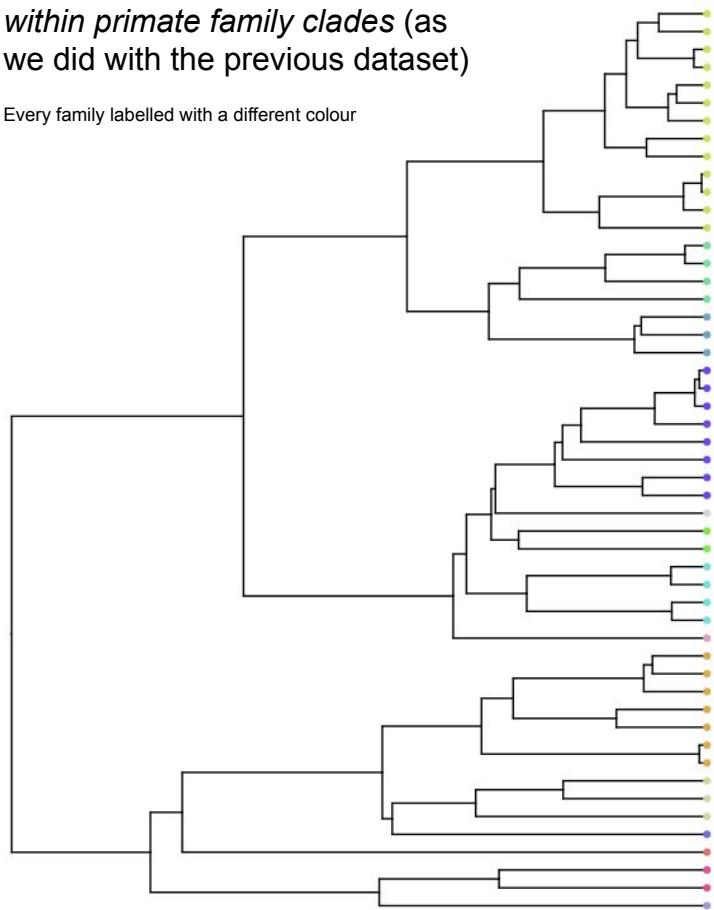
We use a new dataset, with includes only species with >10 observations (more reliability, less diversity of species).

*so there's lots more to look at!



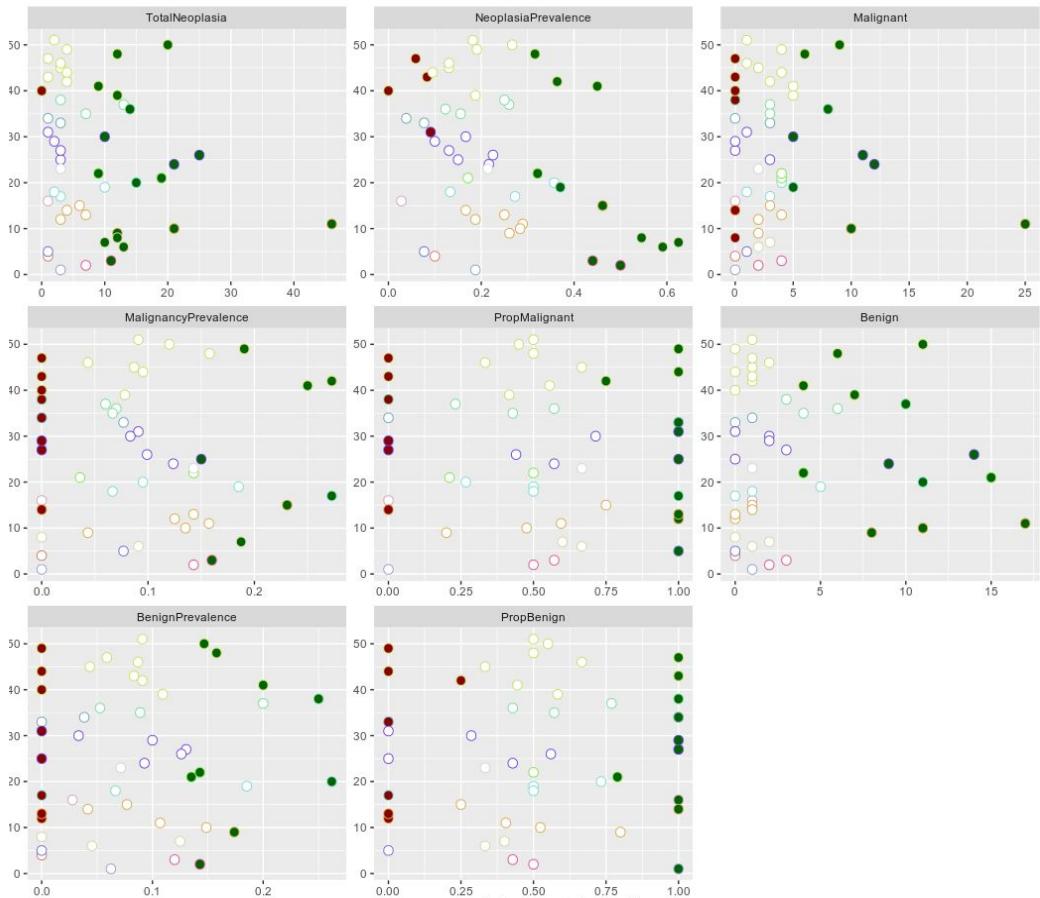
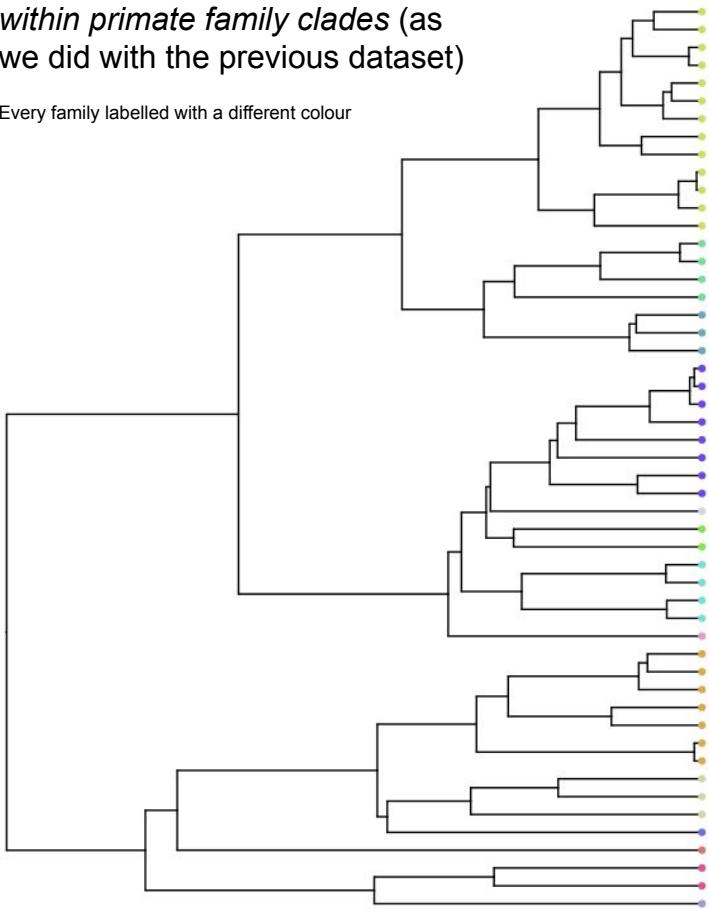
We observe **extreme species**
within primate family clades (as
we did with the previous dataset)

Every family labelled with a different colour



We observe **extreme species**
within primate family clades (as
we did with the previous dataset)

Every family labelled with a different colour



Data quality: a new balance

With the new filtering at the species level, there are two different impacts:

- ↓ Reduction at the species-level **accuracy**. Representative species from the same genus are selected.
- ↑ Improvement of the trait dataset **quality**, with more reliable measurements as mean species values.



***10 measurements x
species,***
balanced trade off

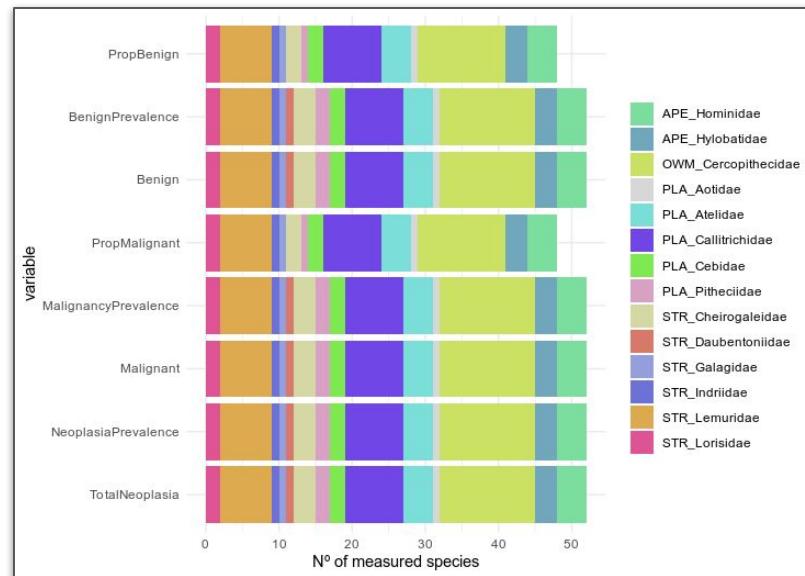
Data quality: a new balance

With the new filtering at the species level, there are two different impacts:

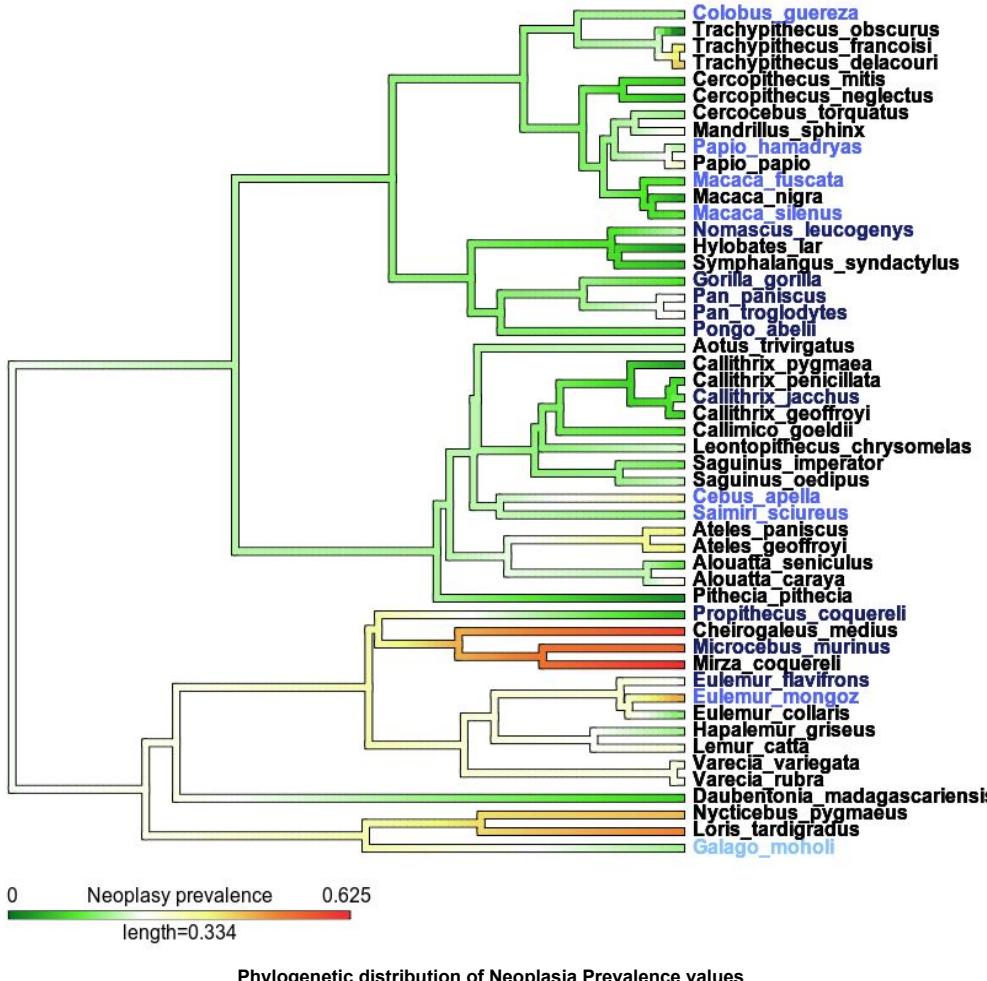
↓ Reduction at the species-level **accuracy**. Representative species from the same genus are selected.

↑ Improvement of the trait dataset **quality**, with more reliable measurements as mean species values.

52 primate species
14 primate families



Available Genomes



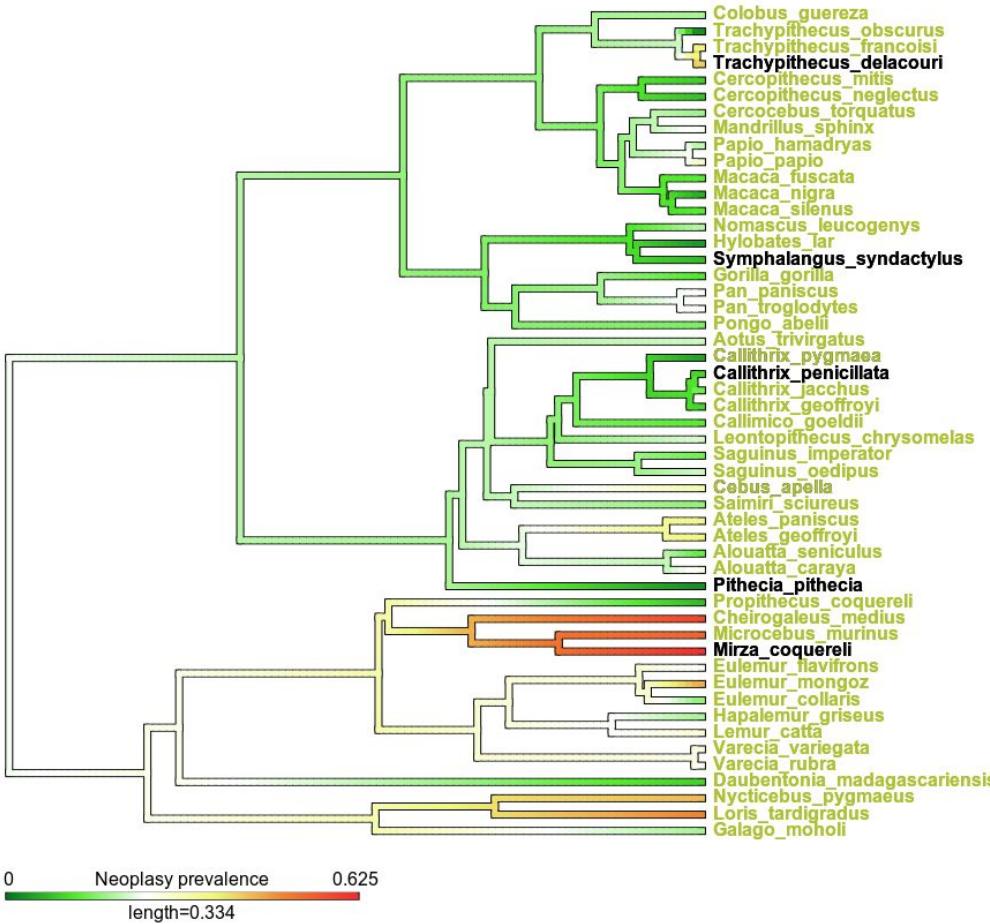
ONLY* 9 species with available UCSC genomes

7 species with genomic data for closest genus representative

1 species with genomic data for closest family representative

*so we are only scratching the surface here!

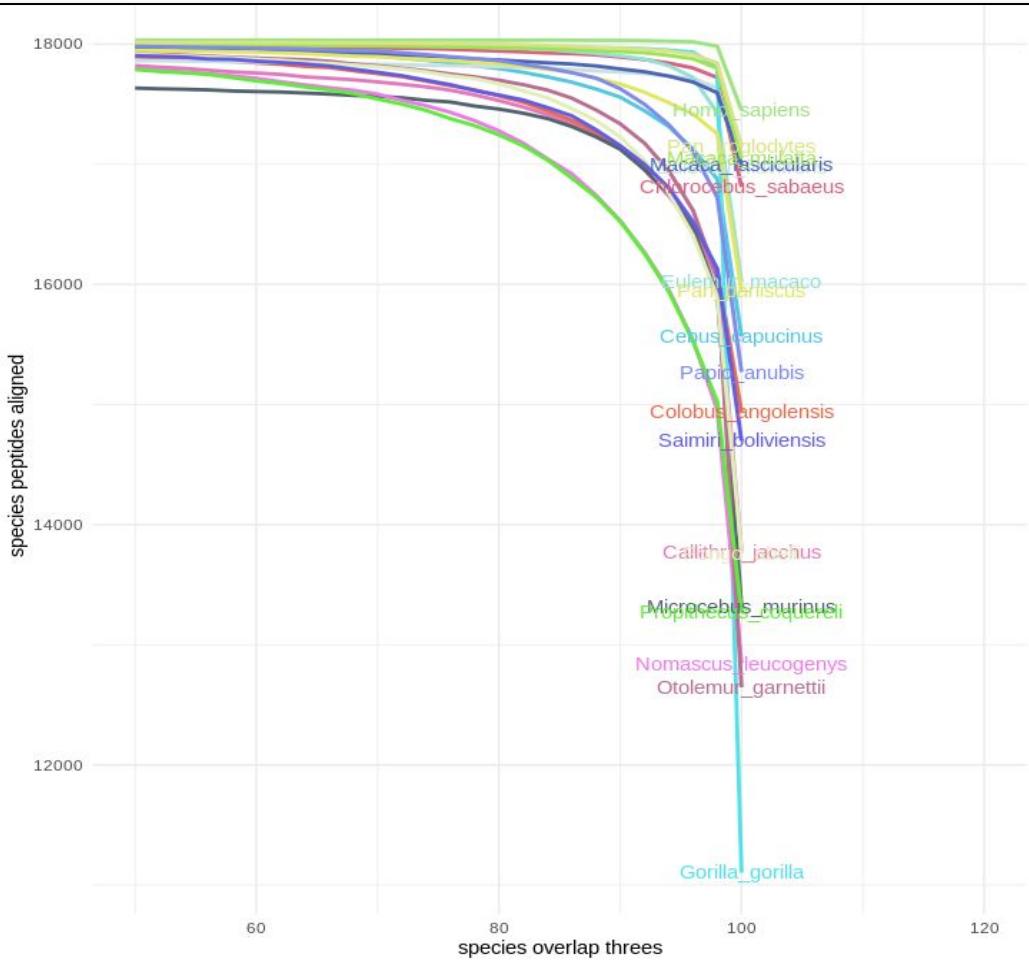
Eventually, we will be
able to add...



~37 species with genomic data for
genomes coming from the
Primate Sequencing Initiative Consortium
(PSIC)

illumina

Final genomic dataset



The analyzed dataset is formed by **16,343 best quality transcript** alignments representing a human gene.

This is our reference genomic dataset for trait analysis.

Discovery of Convergent Amino Acid Substitutions (CAAS)

Discovery of candidate CAAS, focusing on two scenarios, so geared to detecting AA changes linked to Phenotype Group A (e.g., higher neoplasia prevalences)

Scenario 1

Phenotype group A

Species 1 ---- **P S T V Q L M P**

Species 2 ---- **P S T V Q L M P**

Species 3 ---- **P S T V Q L M P**

Phenotype group B

Species 4 ---- **P Q T V Q - M P**

Species 5 ---- **P Q T V Q L M P**

Species 6 ---- **P Q T - Q L M P**

Scenario 2

Phenotype group A

Species 1 ---- **P S T V Q L M P**

Species 2 ---- **P S T V Q L M P**

Species 3 ---- **P S T V Q L M P**

Phenotype group B

Species 4 ---- **P S R V Q - M P**

Species 5 ---- **P S P V Q L M P**

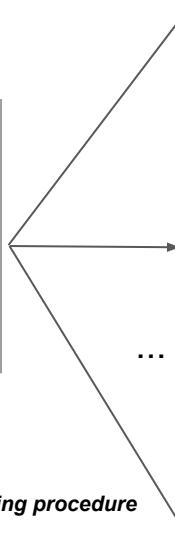
Species 6 ---- **P S M - Q L M P**

Internal Validation of Convergent Amino Acid Substitutions (CAAS)

Phylogeny-guided resampling validation

e.g.	Top species	Bottom species
Trait A	<i>Lemur catta</i> <i>Cercopithecus densus</i>	<i>Erythrocebus patas</i> <i>Macaca arctoides</i> <i>Cheirogaleus major</i> <i>Pan paniscus</i> <i>Propithecus coquereli</i> <i>Saimiri boliviensis</i> <i>Galagooides demidovii</i>

Observed family proportions



*Each color represents a distinct primate family
Primate family proportions from the observed groups are kept in the resampling procedure*

The larger the clade depth, the larger the power to validate AA positions associated to the trait

<u>Resampling 1</u>	Top species	Bottom species
Trait A	<i>Eulemur rufifrons</i> <i>Macaca nigra</i>	<i>Macaca fuscata</i> <i>Colobus angolensis</i> <i>Cheirogaleus medius</i> <i>Pan troglodytes</i> <i>Indri indri</i> <i>Saimiri sciureus</i> <i>Otolemur garnettii</i>
<u>Resampling 2</u>	Top species	Bottom species
Trait A	<i>Eulemur collaris</i> <i>Piliocolobus badius</i>	<i>Presbytis comata</i> <i>Semnopithecus entellus</i> <i>Microcebus murinus</i> <i>Pan paniscus</i> <i>Propithecus diadema</i> <i>Cebus albifrons</i> <i>Galago senegalensis</i>
<u>Resampling N</u>	Top species	Bottom species
Trait A	<i>Eulemur macaco</i> <i>Macaca fuscata</i>	<i>Trachypithecus phayrei</i> <i>Papio papio</i> <i>Mirza zaza</i> <i>Gorilla gorilla</i> <i>Avahi laniger</i> <i>Cebus olivaceus</i> <i>Galago senegalensis</i>

Convergent Amino Acid Substitutions (CAAS)

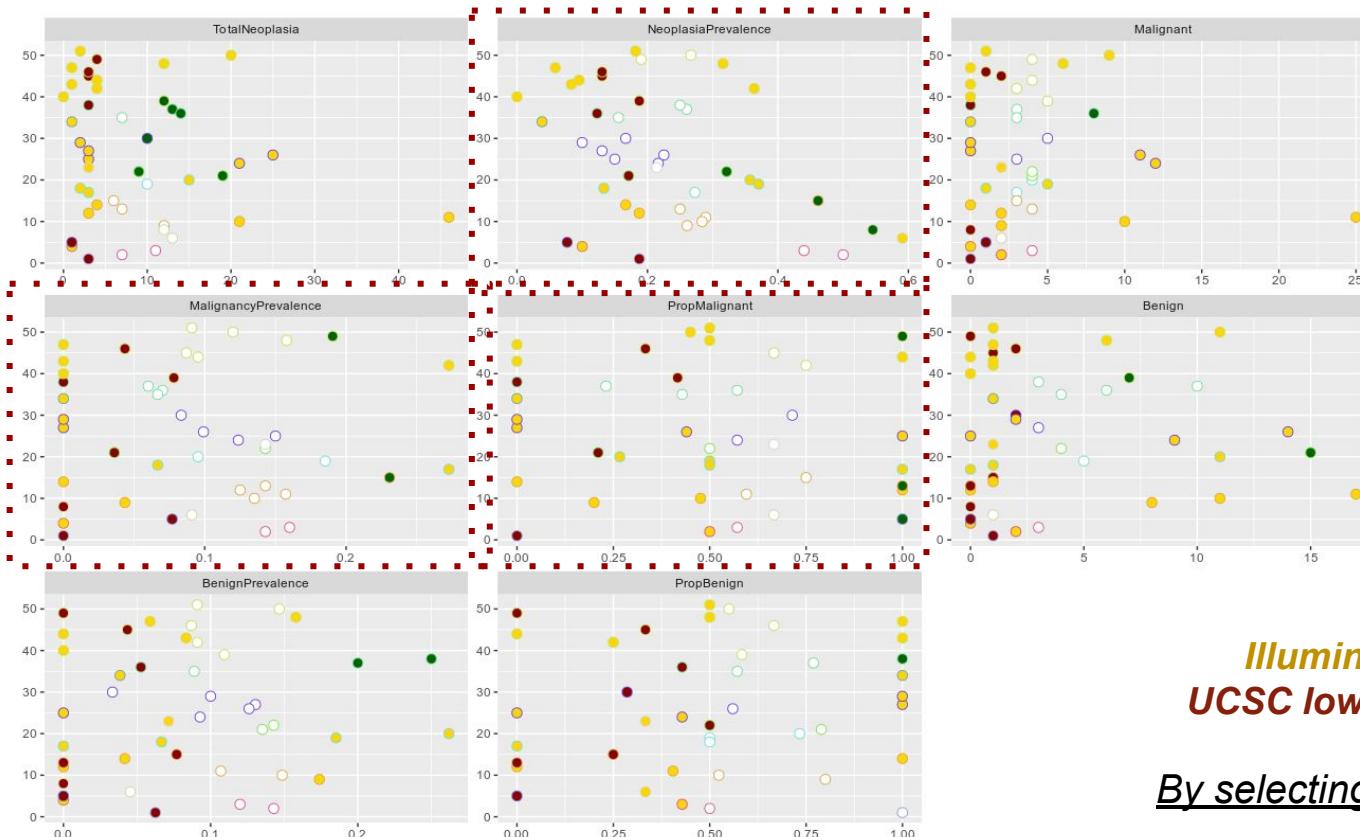
Phylogeny tip order



Trait value

Convergent Amino Acid Substitutions (CAAS)

Phylogeny tip order



*Illumina extreme species
UCSC low-extreme/top-extreme*

By selecting species with ~1 MAD*

*Median Absolute Deviation (from the median both of the family and the global trait distribution)

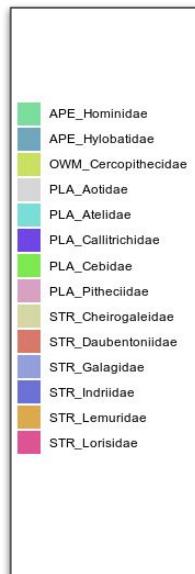
Convergent Amino Acid Substitutions (CAAS)

Malignancy measurements

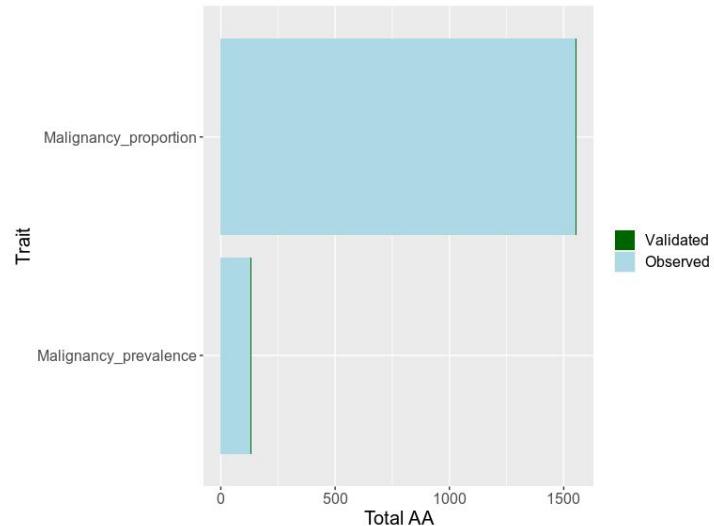
	<i>Top species</i>	<i>Bottom species</i>
Malignancy prevalence	Eulemur mongoz Papio hamadryas	Colobus guereza Macaca silenus Microcebus murinus Pan paniscus Propithecus coquerelii Samiri sciureus Galago moholi
Malignancy proportion	Eulemur flavifrons Papio hamadryas Propithecus coquerelii	Colobus guereza Macaca silenus Pan paniscus Samiri sciureus Galago moholi

1,688 Putative CAAS discovered (134 Malignancy prevalence + 1,554 Malignancy proportion)

703 genes harboring CAAS out of 16,343 alignments (104 Malignancy prevalence + 599 Malignancy proportion)

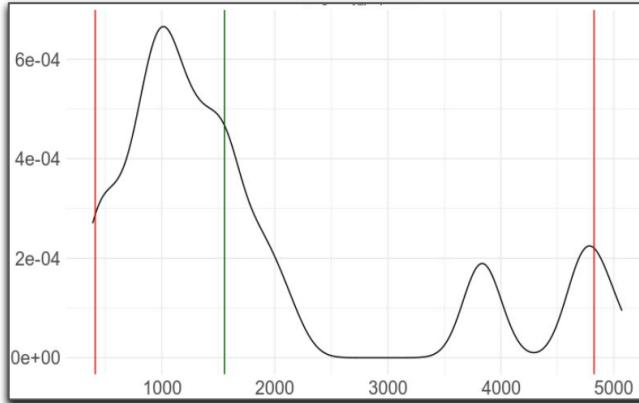


AA discovery + individual validation

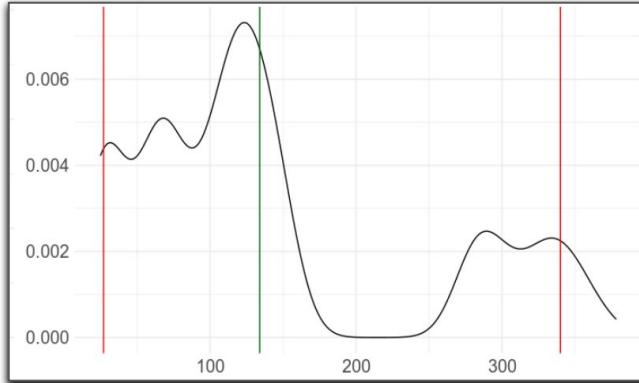


Convergent Amino Acid Substitutions (CAAS)

Malignancy proportion



Malignancy prevalence



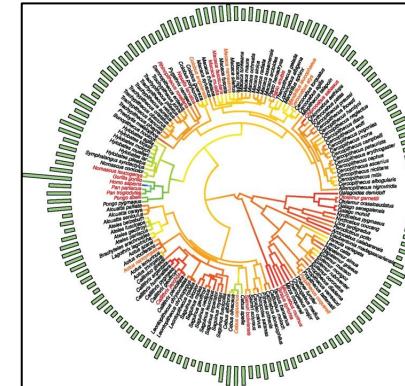
Total n° AA

Amino Acid association excess

Observed nº AA discovered
<0.05 significance area

No excess observed

Even if lack of significant excess due to lack of power (current
of available species aligned), associations found may still be
of relevance



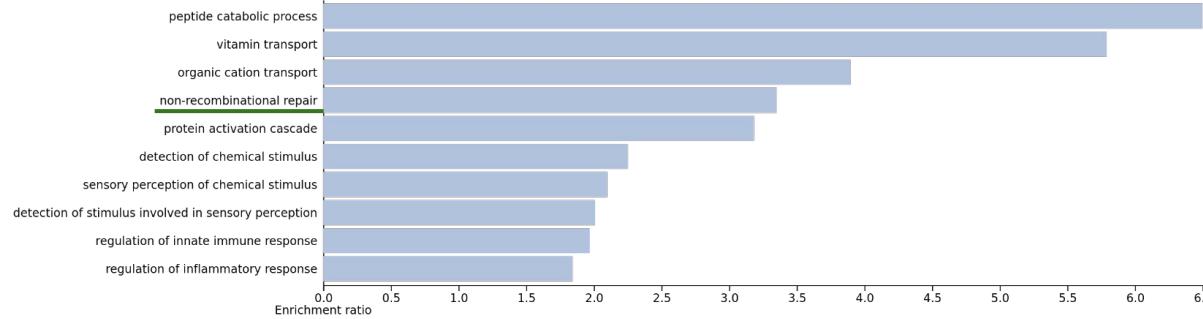
Muntané et al. (2018). Biological processes modulating longevity across primates: A phylogenetic genome-phenome analysis. *Molecular Biology and Evolution*

Convergent Amino Acid Substitutions (CAAS)

FDR ≤ 0.05

FDR > 0.05

Biological process



In-silico functional analysis

GO enrichment analysis

WebGestalt

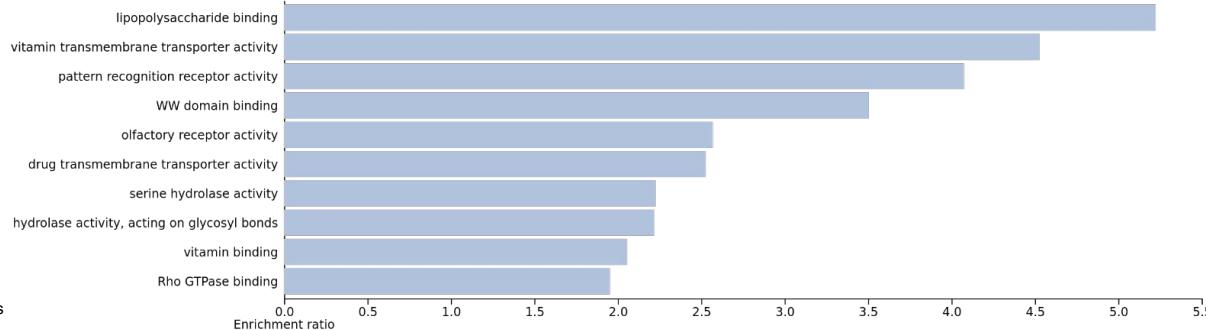
<http://www.webgestalt.org/>

(Analysis performed upon Discovered CAAS)

Reference set to 16,343 genes

Molecular function

FDR ≤ 0.05 FDR > 0.05



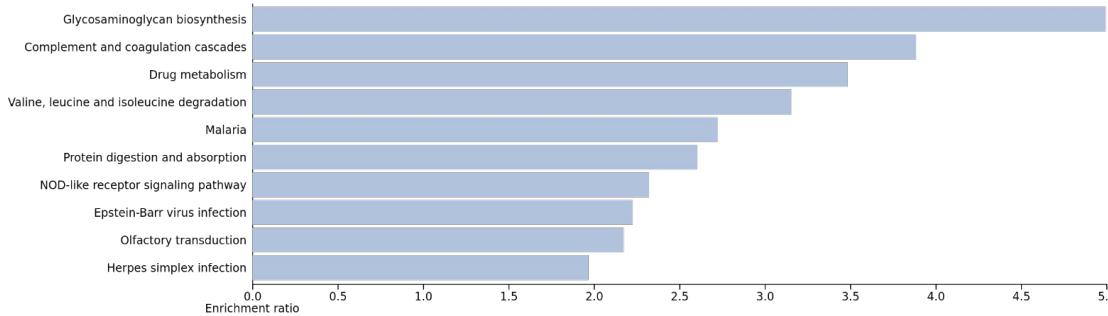
*in green, enrichment categories with several neoplastic genesets

Convergent Amino Acid Substitutions (CAAS)

FDR ≤ 0.05

FDR > 0.05

KEGG 2021: Biological pathways



In-silico functional analysis

Pathway and disease enrichment analysis

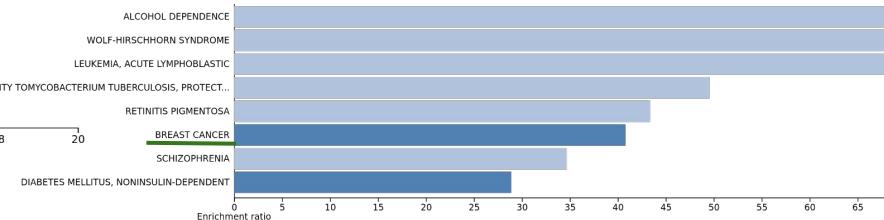
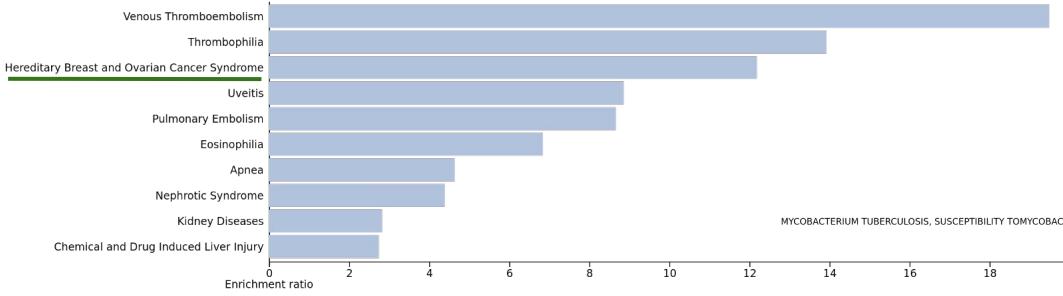
WebGestalt

<http://www.webgestalt.org/>

(Analysis performed upon Discovered CAAS)

Reference set to 16,343 genes

Disgenet & OMIM diseases DB



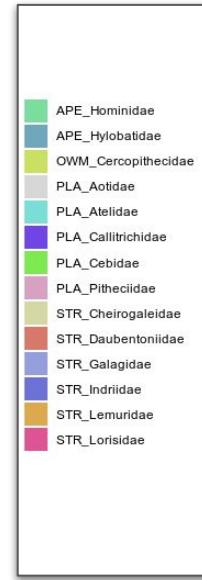
*in green, enrichment categories with several neoplastic genesets

Convergent Amino Acid Substitutions (CAAS)

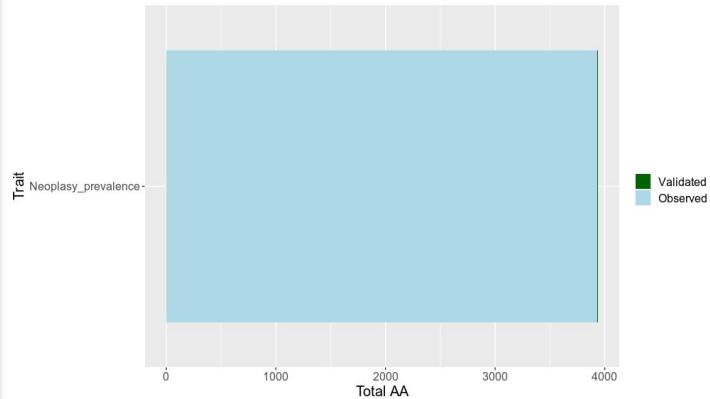
Neoplasia measurements

	Top species	Bottom species
Neoplasia prevalence	<i>Cebus apella</i> <i>Eulemur mongoz</i> <i>Microcebus murinus</i>	<i>Colobus guereza</i> <i>Macaca fuscata</i> <i>Macaca silenus</i> <i>Gorilla gorilla</i> <i>Propithecus coquereli</i> <i>Samiri sciureus</i> <i>Galago moholi</i>

2,261 gene discoveries out of 16,343 alignments
3,939 AA discovered

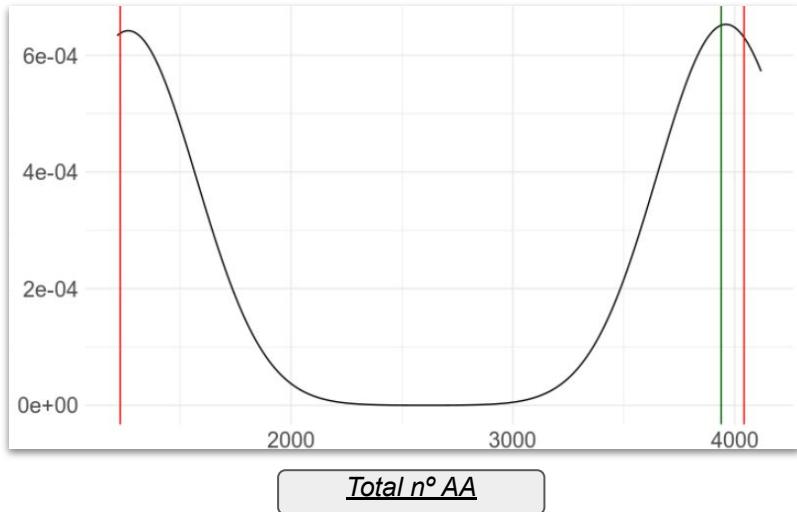


AA discovery + individual validation



Convergent Amino Acid Substitutions (CAAS)

Neoplasia prevalence



Amino Acid association excess

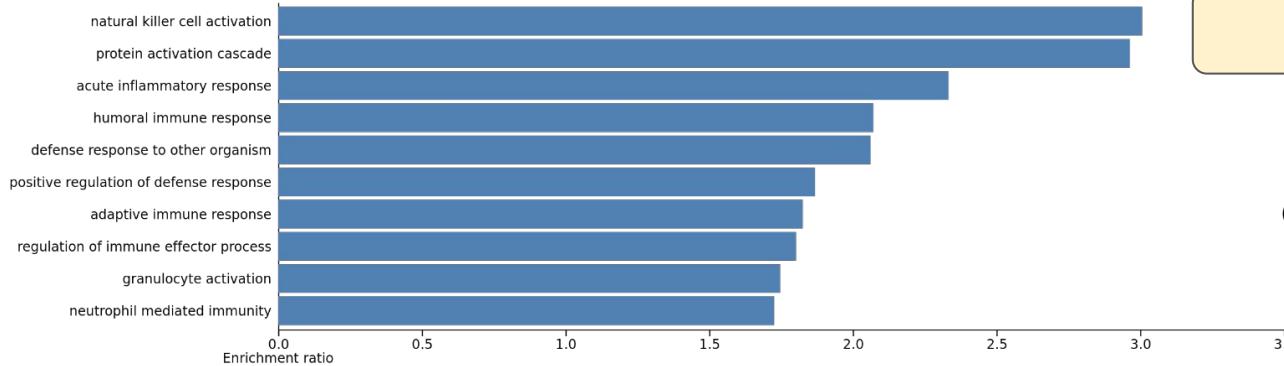
*Observed n° aa discovered
<0.05 significance area*

Convergent Amino Acid Substitutions (CAAS)

FDR ≤ 0.05

FDR > 0.05

Biological process



In-silico functional analysis

GO enrichment analysis

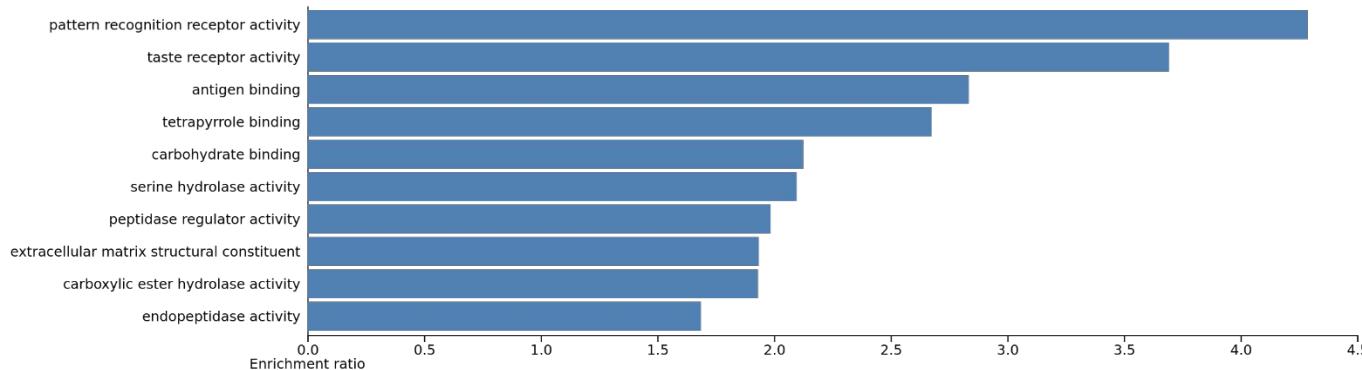
WebGestalt

<http://www.webgestalt.org/>

(Analysis performed upon Discovered CAAS)

Reference set to 16,343 genes

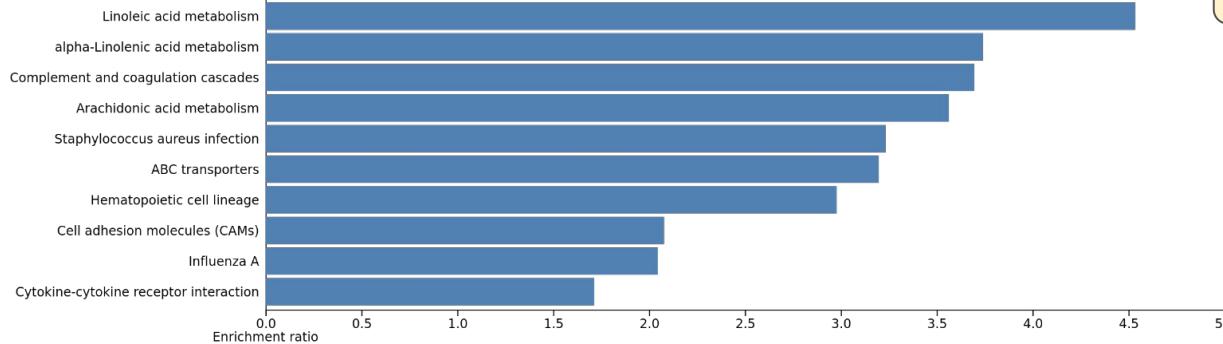
Molecular function



Convergent Amino Acid Substitutions (CAAS)

FDR ≤ 0.05 FDR > 0.05

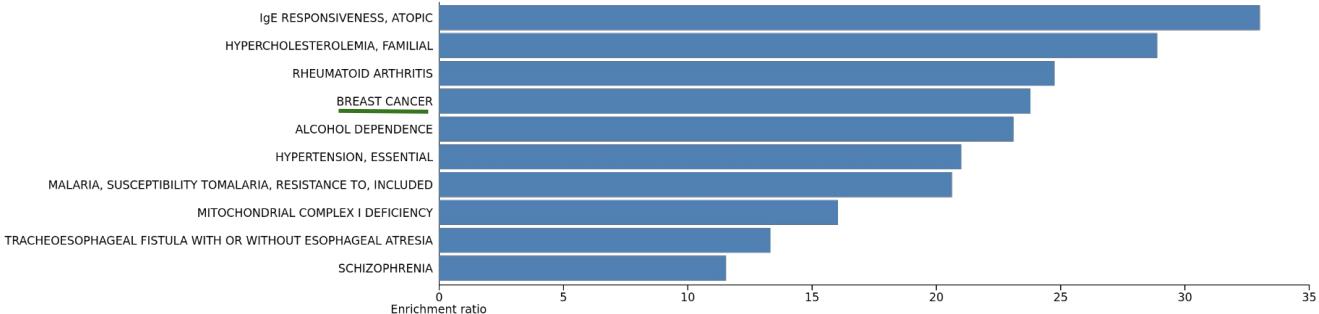
KEGG 2021: Biological pathways



In-silico functional analysis

Pathway and disease enrichment analysis
WebGestalt
<http://www.webgestalt.org/>
(Analysis performed upon Discovered CAAS)
Reference set to 16,343 genes

OMIM diseases



*in green, enrichment categories with several neoplastic genesets

Convergent Amino Acid Substitutions (CAAS)

Table of top 10 significant p-values and q-values for Jensen DISEASES

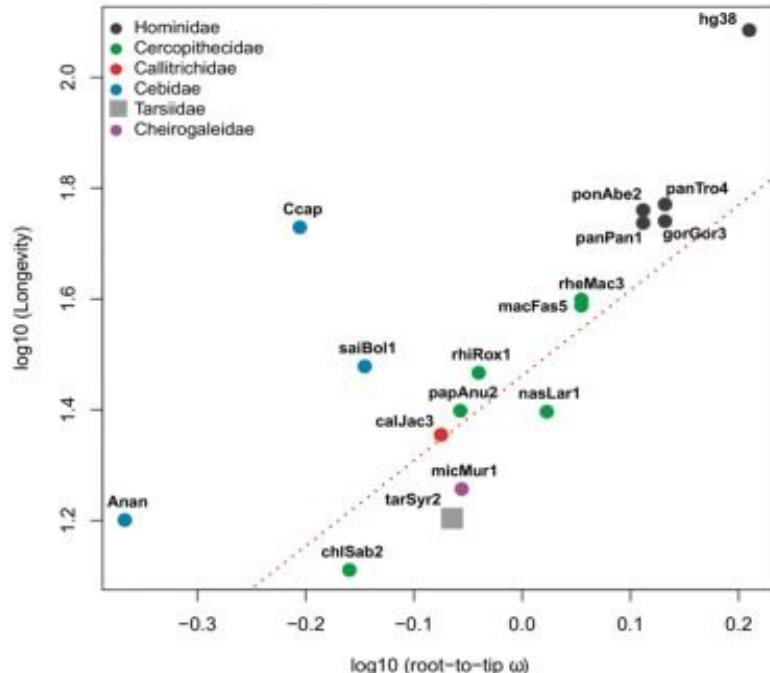
term	p-value	q-value	overlap_genes
Kidney cancer	7.482258e-22	9.412681e-19	[JSP6NL, FRMPD2, SLCO6A1, ACCS, ABCA12, SLCA44, BAC1H, ABCA13, ZDBF2, KIAA1109, ZFYVE26, NCKX4P5, NUP210, CACNA2D1, WDR72, HSPG2, CPED1, LNBD1, ACE2, FLG2, ABCA10, ZFYVE16, NUP210L, TEP1, ARMC3, HECW1, MAP1A, FREM2, ENPEP, ADAMTS12, RNFL213, C6, EPB4112, CEP70, PCDHGA12, FAM135B, PCDH9B, TCN2, FASN, FAT2, LITD1, D2IP3, PIGR, SEPF2, DENND1A, FMN2, ABCC11, ABCC12, MRC2, JPH3, CHL1, CNP, GPRAS1, NINL, ZNF483, SHPRH, WDR36, FCR5L, FCR5, RPKA, TEK14, NCOR1, EXOC4, PLCB1, CRB1, NBAS, SAM9, SYNPO2, PRUNE1, ABCB11, AXNDND1, PTGS1, ADAM29, CSMD1, WDR17, PARP1, UTP6, ATAD2, PRSS35, CENP, ZNF1X1, CENP, SLC04C1, ZNF217, ZNF335, ZNF454, ...]
Skin cancer	3.154536e-19	1.984203e-16	[FCGBP, ADAMDEC1, RGS1L, ABCA12, GIMAP2, TGM3, PGDGRA, HGF, ANK2, ANK3, PCDH4A10, CPED1, FLG2, SLCA94, KCNQ3, COL4A6, ADAM7, ALPK2, ZNF831, PLA2R1, MXRA5, FREM2, DSC3, MACF1, FBNI, ENPEP, SLC4524, NOTCH1, IGSF1, CD1C, CD1B, PKHD1, C6, SLN, LRPF2, MOV10L1, APOB, MORC1, TRPM6, FAM83B, TMEM132D, MSR1, ABCA6, ABCA3, ARHGAP29, PRGA, TDROD, ISX, CYP2C9, ADCY10, MARCO, FAM135B, PTPRC, ALB, COL5A2, COL9A1, ZNF538, LTD1, ATM, SPAG17, COL14A1, LRRK2, STAB2, CHODC, FMN2, EFCAB6, PTPRB, ABC112, SYNE1, CHL1, RNF17, HYDN, AOX1, ADAMTS7, DSP, MYBP1, VPS13C, VPS13A, TEK14, OSMR, MMPI8, SPTB, ADAMTS20, VCAN, CATSPERB, DSG1, DS3, PLCBL1, DS64, LCT, CYLC2, TRA, DNAH3, CRB1, C15, LAMA1, SYNPO2, DNAH5, DNAH7, ...]
Melanoma	9.885882e-17	4.145480e-14	[TSK5, FCGBP, ADAMDEC1, RGS1L, ABCA12, GIMAP2, TGM3, ANK2, ANK3, CPED1, FLG2, CNB2D, SLCA94, ZNF318, KCNQ3, COL4A6, DPEP1, ADAM7, R0R2, ALPK2, ZNF831, MXRA5, FREM2, DSC3, ZNF276, MACF1, FBNI, ENPEP, SLC4524, IGSF1, CYP2C9, CD1C, CD1B, IGSF9, PKHD1, C6, SLN, LRPF2, MOV10L1, APOB, MORC1, TRPM6, FAM83B, MSR1, DNHA10, ABCA6, SPAG16, PGR, SPAG17, COL14A1, LRRK2, CRISP2, STAB2, FMN2, EFCAB6, SYNE1, CHL1, HYDN, AOX1, ANXA9, ADAMTS7, DSP, FCR5L, VPS13C, VPS13A, OSMR, MMPI8, SPTB, ADAMTS20, VCAN, CATSPERB, CNGB3, DSg1, DS3, PLCBL1, DS64, LCT, CYLC2, TRA, DNAH3, CRB1, C15, LAMA1, SYNPO2, DNAH5, DNAH7, ...]
Carcinoma	1.513483e-16	4.759904e-14	[TSK5, USP6NL, PGYRP3, C4BPA, ACCS, ZDBF2, L4L1, KIAA1109, ZFYVE26, SCP2, LIP1, LIP1, C3A1L, UBASH3A, SAMD11, ZNF17, GBP7, ZNF218, CACNA2D1, STAR9D, HSPG2, CNB2D, ACE2, SLC5A7, FLG2, CNB2D, ZFVE26, NUP210L, TMEM79, TMEM126A, SPARC1, NBRK, ENTHD1, P0A1, ZNF276, URH2, C2RET3, RETSAT, AASDH, CSORF49, KTRAP27-1, PL2G23, TEK26, ZNF23, HCV1, NEDD1, C6, CYP2B6, HLC5, EPB4112, AGMO, C9, MOV10L1, PIP5K1A, HAO1, LT14H, EXPHA, JAG2, ABCA1, CTCL, ABCA2, CSORF34, COL24A1, ABCA1, ABCA3, FUCA2, L3L1RA, ABCA1, EFCAK2, L1L2RA1, SETX, FAM111A, PCDHGA11, CYP2C9, PTPRB1, PCDHGA12, FAM135B, ZFAF1, HRC, C11ORF16, FAT2, LTD1, C19ORF44, D2P1, D2P2, NOX1, GMFN, FERM73, ...]
Liver cancer	4.187366e-14	1.053541e-11	[JSP6NL, MKB67, ABCA12, SLCA94, LIP1, KIAA1109, EXP4, ADGB, HGF, STAR0D, TSC2, ANK2, VMD2, ANK3, HSPG2, SLC9A7, OBSN, HECW1, KINQ3, ZNF831, PLA2R1, MXRA5, FREM2, MACF1, NOTCH1, IGSF1, BOD1L1, ADAMTS12, PKHD1, SPTA1, RNF123, P0D2Z, MAP2, SLI, CLCA2, HIVEP1, TRPM5, APOB, TRPM6, MN1, FAM83B, ABCA1, TMEM132D, DNHA12, DNHA10, COL24A1, SYNRG, NEB, ZFHX4, SETX, GSF10, SOK1, TF, GONAL, PTPRC, FASN, WNKC, GOLGB1, ALB, FAT2, ZNF536, ATM, BCRL1, NEDD1, C6, CYP2B6, MN1, SCD22, SPEF2, MCP1, SYCP2, MAST4, DENND1A, MCMB1, STAB2, CHD6, FMN2, BRCAL1, SYNE1, SYNE1, JPHS, HCN2, LIP1, HEZ2, P0A1, SPTA1, ADAM29, P0D2Z, CPAMOB, PLXNA2, APOB, PCNT, CSMD1, MU4, MDN1, IMPG1, TMEM132D, SPEN, GALNT5, ANHAK2, LAMB4, NEB, TDID6, USH2A, ZFHX4, SDK1, FRAS1, MYO15A, MAP3K19, GOLGB1, COL7A1, NEK10, FAT2, LT17, ABCA1, BCRL1, BCR, BCRL2, BCRL3, BCRL4, BCRL5, BCRL6, BCRL7, BCRL8, BCRL9, BCRL10, BCRL11, BCRL12, BCRL13, BCRL14, BCRL15, BCRL16, BCRL17, BCRL18, BCRL19, BCRL20, BCRL21, BCRL22, BCRL23, BCRL24, BCRL25, BCRL26, BCRL27, BCRL28, BCRL29, BCRL30, BCRL31, BCRL32, BCRL33, BCRL34, BCRL35, BCRL36, BCRL37, BCRL38, BCRL39, BCRL40, BCRL41, BCRL42, BCRL43, BCRL44, BCRL45, BCRL46, BCRL47, BCRL48, BCRL49, BCRL50, BCRL51, BCRL52, BCRL53, BCRL54, BCRL55, BCRL56, BCRL57, BCRL58, BCRL59, BCRL60, BCRL61, BCRL62, BCRL63, BCRL64, BCRL65, BCRL66, BCRL67, BCRL68, BCRL69, BCRL70, BCRL71, BCRL72, BCRL73, BCRL74, BCRL75, BCRL76, BCRL77, BCRL78, BCRL79, BCRL80, BCRL81, BCRL82, BCRL83, BCRL84, BCRL85, BCRL86, BCRL87, BCRL88, BCRL89, BCRL90, BCRL91, BCRL92, BCRL93, BCRL94, BCRL95, BCRL96, BCRL97, BCRL98, BCRL99, BCRL100, BCRL101, BCRL102, BCRL103, BCRL104, BCRL105, BCRL106, BCRL107, BCRL108, BCRL109, BCRL110, BCRL111, BCRL112, BCRL113, BCRL114, BCRL115, BCRL116, BCRL117, BCRL118, BCRL119, BCRL120, BCRL121, BCRL122, BCRL123, BCRL124, BCRL125, BCRL126, BCRL127, BCRL128, BCRL129, BCRL130, BCRL131, BCRL132, BCRL133, BCRL134, BCRL135, BCRL136, BCRL137, BCRL138, BCRL139, BCRL140, BCRL141, BCRL142, BCRL143, BCRL144, BCRL145, BCRL146, BCRL147, BCRL148, BCRL149, BCRL150, BCRL151, BCRL152, BCRL153, BCRL154, BCRL155, BCRL156, BCRL157, BCRL158, BCRL159, BCRL160, BCRL161, BCRL162, BCRL163, BCRL164, BCRL165, BCRL166, BCRL167, BCRL168, BCRL169, BCRL170, BCRL171, BCRL172, BCRL173, BCRL174, BCRL175, BCRL176, BCRL177, BCRL178, BCRL179, BCRL180, BCRL181, BCRL182, BCRL183, BCRL184, BCRL185, BCRL186, BCRL187, BCRL188, BCRL189, BCRL190, BCRL191, BCRL192, BCRL193, BCRL194, BCRL195, BCRL196, BCRL197, BCRL198, BCRL199, BCRL200, BCRL201, BCRL202, BCRL203, BCRL204, BCRL205, BCRL206, BCRL207, BCRL208, BCRL209, BCRL210, BCRL211, BCRL212, BCRL213, BCRL214, BCRL215, BCRL216, BCRL217, BCRL218, BCRL219, BCRL220, BCRL221, BCRL222, BCRL223, BCRL224, BCRL225, BCRL226, BCRL227, BCRL228, BCRL229, BCRL230, BCRL231, BCRL232, BCRL233, BCRL234, BCRL235, BCRL236, BCRL237, BCRL238, BCRL239, BCRL240, BCRL241, BCRL242, BCRL243, BCRL244, BCRL245, BCRL246, BCRL247, BCRL248, BCRL249, BCRL250, BCRL251, BCRL252, BCRL253, BCRL254, BCRL255, BCRL256, BCRL257, BCRL258, BCRL259, BCRL260, BCRL261, BCRL262, BCRL263, BCRL264, BCRL265, BCRL266, BCRL267, BCRL268, BCRL269, BCRL270, BCRL271, BCRL272, BCRL273, BCRL274, BCRL275, BCRL276, BCRL277, BCRL278, BCRL279, BCRL280, BCRL281, BCRL282, BCRL283, BCRL284, BCRL285, BCRL286, BCRL287, BCRL288, BCRL289, BCRL290, BCRL291, BCRL292, BCRL293, BCRL294, BCRL295, BCRL296, BCRL297, BCRL298, BCRL299, BCRL200, BCRL201, BCRL202, BCRL203, BCRL204, BCRL205, BCRL206, BCRL207, BCRL208, BCRL209, BCRL2010, BCRL2011, BCRL2012, BCRL2013, BCRL2014, BCRL2015, BCRL2016, BCRL2017, BCRL2018, BCRL2019, BCRL2020, BCRL2021, BCRL2022, BCRL2023, BCRL2024, BCRL2025, BCRL2026, BCRL2027, BCRL2028, BCRL2029, BCRL2030, BCRL2031, BCRL2032, BCRL2033, BCRL2034, BCRL2035, BCRL2036, BCRL2037, BCRL2038, BCRL2039, BCRL2040, BCRL2041, BCRL2042, BCRL2043, BCRL2044, BCRL2045, BCRL2046, BCRL2047, BCRL2048, BCRL2049, BCRL2050, BCRL2051, BCRL2052, BCRL2053, BCRL2054, BCRL2055, BCRL2056, BCRL2057, BCRL2058, BCRL2059, BCRL2060, BCRL2061, BCRL2062, BCRL2063, BCRL2064, BCRL2065, BCRL2066, BCRL2067, BCRL2068, BCRL2069, BCRL2070, BCRL2071, BCRL2072, BCRL2073, BCRL2074, BCRL2075, BCRL2076, BCRL2077, BCRL2078, BCRL2079, BCRL2080, BCRL2081, BCRL2082, BCRL2083, BCRL2084, BCRL2085, BCRL2086, BCRL2087, BCRL2088, BCRL2089, BCRL2090, BCRL2091, BCRL2092, BCRL2093, BCRL2094, BCRL2095, BCRL2096, BCRL2097, BCRL2098, BCRL2099, BCRL20100, BCRL20101, BCRL20102, BCRL20103, BCRL20104, BCRL20105, BCRL20106, BCRL20107, BCRL20108, BCRL20109, BCRL20110, BCRL20111, BCRL20112, BCRL20113, BCRL20114, BCRL20115, BCRL20116, BCRL20117, BCRL20118, BCRL20119, BCRL20120, BCRL20121, BCRL20122, BCRL20123, BCRL20124, BCRL20125, BCRL20126, BCRL20127, BCRL20128, BCRL20129, BCRL20130, BCRL20131, BCRL20132, BCRL20133, BCRL20134, BCRL20135, BCRL20136, BCRL20137, BCRL20138, BCRL20139, BCRL20140, BCRL20141, BCRL20142, BCRL20143, BCRL20144, BCRL20145, BCRL20146, BCRL20147, BCRL20148, BCRL20149, BCRL20150, BCRL20151, BCRL20152, BCRL20153, BCRL20154, BCRL20155, BCRL20156, BCRL20157, BCRL20158, BCRL20159, BCRL20160, BCRL20161, BCRL20162, BCRL20163, BCRL20164, BCRL20165, BCRL20166, BCRL20167, BCRL20168, BCRL20169, BCRL20170, BCRL20171, BCRL20172, BCRL20173, BCRL20174, BCRL20175, BCRL20176, BCRL20177, BCRL20178, BCRL20179, BCRL20180, BCRL20181, BCRL20182, BCRL20183, BCRL20184, BCRL20185, BCRL20186, BCRL20187, BCRL20188, BCRL20189, BCRL201800, BCRL201801, BCRL201802, BCRL201803, BCRL201804, BCRL201805, BCRL201806, BCRL201807, BCRL201808, BCRL201809, BCRL2018010, BCRL2018011, BCRL2018012, BCRL2018013, BCRL2018014, BCRL2018015, BCRL2018016, BCRL2018017, BCRL2018018, BCRL2018019, BCRL20180100, BCRL20180101, BCRL20180102, BCRL20180103, BCRL20180104, BCRL20180105, BCRL20180106, BCRL20180107, BCRL20180108, BCRL20180109, BCRL20180110, BCRL20180111, BCRL20180112, BCRL20180113, BCRL20180114, BCRL20180115, BCRL20180116, BCRL20180117, BCRL20180118, BCRL20180119, BCRL201801100, BCRL201801101, BCRL201801102, BCRL201801103, BCRL201801104, BCRL201801105, BCRL201801106, BCRL201801107, BCRL201801108, BCRL201801109, BCRL201801110, BCRL201801111, BCRL201801112, BCRL201801113, BCRL201801114, BCRL201801115, BCRL201801116, BCRL201801117, BCRL201801118, BCRL201801119, BCRL2018011100, BCRL2018011101, BCRL2018011102, BCRL2018011103, BCRL2018011104, BCRL2018011105, BCRL2018011106, BCRL2018011107, BCRL2018011108, BCRL2018011109, BCRL2018011110, BCRL2018011111, BCRL2018011112, BCRL2018011113, BCRL2018011114, BCRL2018011115, BCRL2018011116, BCRL2018011117, BCRL2018011118, BCRL2018011119, BCRL20180111100, BCRL20180111101, BCRL20180111102, BCRL20180111103, BCRL20180111104, BCRL20180111105, BCRL20180111106, BCRL20180111107, BCRL20180111108, BCRL20180111109, BCRL20180111110, BCRL20180111111, BCRL20180111112, BCRL20180111113, BCRL20180111114, BCRL20180111115, BCRL20180111116, BCRL20180111117, BCRL20180111118, BCRL20180111119, BCRL201801111100, BCRL201801111101, BCRL201801111102, BCRL201801111103, BCRL201801111104, BCRL201801111105, BCRL201801111106, BCRL201801111107, BCRL201801111108, BCRL201801111109, BCRL201801111110, BCRL201801111111, BCRL201801111112, BCRL201801111113, BCRL201801111114, BCRL201801111115, BCRL201801111116, BCRL201801111117, BCRL201801111118, BCRL201801111119, BCRL2018011111100, BCRL2018011111101, BCRL2018011111102, BCRL2018011111103, BCRL2018011111104, BCRL2018011111105, BCRL2018011111106, BCRL2018011111107, BCRL2018011111108, BCRL2018011111109, BCRL2018011111110, BCRL2018011111111, BCRL2018011111112, BCRL2018011111113, BCRL2018011111114, BCRL2018011111115, BCRL2018011111116, BCRL2018011111117, BCRL2018011111118, BCRL2018011111119, BCRL20180111111100, BCRL20180111111101, BCRL20180111111102, BCRL20180111111103, BCRL20180111111104, BCRL20180111111105, BCRL20180111111106, BCRL20180111111107, BCRL20180111111108, BCRL20180111111109, BCRL20180111111110, BCRL20180111111111, BCRL20180111111112, BCRL20180111111113, BCRL20180111111114, BCRL20180111111115, BCRL20180111111116, BCRL20180111111117, BCRL20180111111118, BCRL20180111111119, BCRL201801111111100, BCRL201801111111101, BCRL201801111111102, BCRL201801111111103, BCRL201801111111104, BCRL201801111111105, BCRL201801111111106, BCRL201801111111107, BCRL201801111111108, BCRL201801111111109, BCRL201801111111110, BCRL201801111111111, BCRL201801111111112, BCRL201801111111113, BCRL201801111111114, BCRL201801111111115, BCRL201801111111116, BCRL201801111111117, BCRL201801111111118, BCRL201801111111119, BCRL2018011111111100, BCRL2018011111111101, BCRL2018011111111102, BCRL2018011111111103, BCRL2018011111111104, BCRL2018011111111105, BCRL2018011111111106, BCRL2018011111111107, BCRL2018011111111108, BCRL2018011111111109, BCRL2018011111111110, BCRL2018011111111111, BCRL2018011111111112, BCRL2018011111111113, BCRL2018011111111114, BCRL2018011111111115, BCRL2018011111111116, BCRL2018011111111117, BCRL2018011111111118, BCRL2018011111111119, BCRL20180111111111100, BCRL20180111111111101, BCRL20180111111111102, BCRL20180111111111103, BCRL20180111111111104, BCRL20180111111111105, BCRL20180111111111106, BCRL20180111111111107, BCRL20180111111111108, BCRL20180111111111109, BCRL20180111111111110, BCRL20180111111111111, BCRL20180111111111112, BCRL20180111111111113, BCRL20180111111111114, BCRL20180111111111115, BCRL20180111111111116, BCRL20180111111111117, BCRL20180111111111118, BCRL20180111111111119, BCRL201801111111111100, BCRL201801111111111101, BCRL201801111111111102, BCRL201801111111111103, BCRL201801111111111104, BCRL201801111111111105, BCRL201801111111111106, BCRL201801111111111107, BCRL201801111111111108, BCRL201801111111111109, BCRL201801111111111110, BCRL201801111111111111, BCRL201801111111111112, BCRL201801111111111113, BCRL201801111111111114, BCRL201801111111111115, BCRL201801111111111116, BCRL201801111111111117, BCRL201801111111111118, BCRL201801111111111119, BCRL2018011111111111100, BCRL2018011111111111101, BCRL2018011111111111102, BCRL2018011111111111103, BCRL2018011111111111104, BCRL2018011111111111105, BCRL2018011111111111106, BCRL2018011111111111107, BCRL2018011111111111108, BCRL2018011111111111109, BCRL2018011111111111110, BCRL2018011111111111111, BCRL2018011111111111112, BCRL2018011111111111113, BCRL2018011111111111114, BCRL2018011111111111115, BCRL2018011111111111116, BCRL2018011111111111117, BCRL2018011111111111118, BCRL2018011111111111119, BCRL20180111111111111100, BCRL20180111111111111101, BCRL20180111111111111102, BCRL20180111111111111103, BCRL20180111111111111104, BCRL20180111111111111105, BCRL20180111111111111106, BCRL20180111111111111107, BCRL20180111111111111108, BCRL20180111111111111109, BCRL20180111111111111110, BCRL20180111111111111111, BCRL20180111111111111112, BCRL20180111111111111113, BCRL20180111111111111114, BCRL20180111111111111115, BCRL20180111111111111116, BCRL20180111111111111117, BCRL20180111111111111118, BCRL20180111111111111119, BCRL201801111111111111100, BCRL201801111111111111101, BCRL201801111111111111102, BCRL201801111111111111103, BCRL201801111111111111104, BCRL201801111111111111105, BCRL201801111111111111106, BCRL201801111111111111107, BCRL201801111111111111108, BCRL201801111111111111109, BCRL201801111111111111110, BCRL201801111111111111111, BCRL201801111111111111112, BCRL201801111111111111113, BCRL201801111111111111114, BCRL201801111111111111115, BCRL201801111111111111116, BCRL201801111111111111117, BCRL201801111111111111118, BCRL201801111111111111119, BCRL2018011111111111111100, BCRL2018011111111111111101, BCRL2018011111111111111102, BCRL2018011111111111111103, BCRL2018011111111111111104, BCRL2018011111111111111105, BCRL2018011111111111111106, BCRL2018011111111111111107, BCRL2018011111111111111108, BCRL2018011111111111111109, BCRL2018011111111111111110, BCRL2018011111111111111111, BCRL2018011111111111111112, BCRL2018011111111111111113, BCRL2018011111111111111114, BCRL2018011111111111111115, BCRL2018011111111111111116, BCRL2018011111111111111117, BCRL2018011111111111111118, BCRL2018011111111111111119, BCRL20180111111111111111100, BCRL20180111111111111111101, BCRL20180111111111111111102, BCRL20180111111111111111103, BCRL20180111111111111111104, BCRL20180111111111111111105, BCRL20180111111111111111106, BCRL20180111111111111111107, BCRL20180111111111111111108, BCRL20180111111111111111109, BCRL20180111111111111111110, BCRL20180111111111111111111, BCRL2018011111111111111112, BCRL2018011111111111111113, BCRL2018011111111111111114, BCRL2018011111111111111115, BCRL2018011111111111111116, BCRL2018011111111111111117, BCRL2018011111111111111118, BCRL2018011111111111111119, BCRL20180111111111111111100, BCRL20180111111111111111101, BCRL20180111111111111111102, BCRL20180111111111111111103, BCRL20180111111111111111104, BCRL20180111111111111111105, BCRL20180111111111111111106, BCRL20180111111111111111107, BCRL20180111111111111111108, BCRL20180111111111111111109, BCRL20180111111111111111110, BCRL20180111111111111111111, BCRL2018011111111111111112, BCRL2018011111111111111113, BCRL2018011111111111111114, BCRL201801111111111

Summary of CAAS analysis

- **Malignancy related-trait**s represent close selection of extreme primate species, with the exception of *Propithecus coquereli*, in which genes associated with both traits present AAs associated with an increase of the proportion of Malignancy over Benignancy but an absent increase of their Malignancy rate.
- Top cancer-related categories for their genesets are linked to **Ovarian and Breast cancer** with genes such as BRCA1, RAD51D and BARD1 in the top enrichment list.
- Neoplasia prevalence also links to **fatty acid metabolism** and **immunity/inflammatory responses**, with a strong signal in cancer categories for one of the available human disease annotation DB (Jensen diseases).

PGLS-based trait-association analysis

(Phylogenetic Generalised Least-Squares, PGLS)



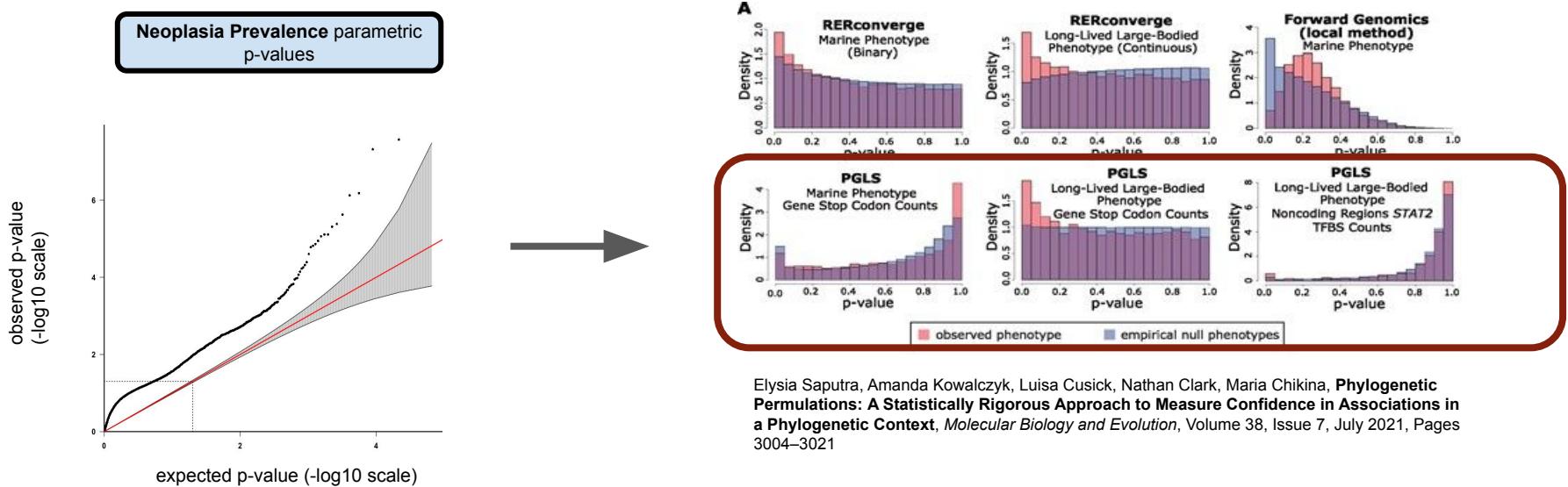
Regression between a genomic variable (in our case, rates of protein evolution, measured as **root-to-tip dN/dS**) and a trait response variable (e.g., **Neoplasia Prevalence**) across a phylogeny, controlling for non-independence of species data points.

- + *gene-trait correlations*
- *gene-trait correlations*

Use of permutations

For a given phylogenetic comparative method (PCM), such as PGLS, permutations allow to:

- explore the space** of null expectation for p-values given a particular association method and a particular genomic dataset
- recalibrate** original p-values obtaining **empirical p-values**, reducing overestimation or underestimation of correlation significance due to PCM particularities

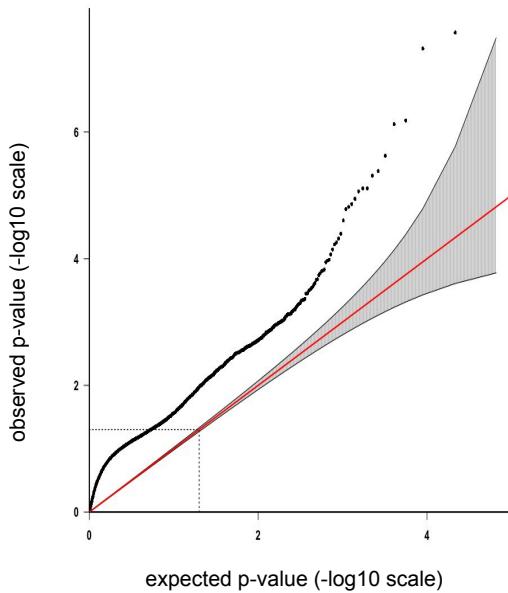


Improvements added since last presentation...

Regarding both, the gene predictor (root-to-tip dN/dS) and the relevant trait (e.g., Neoplasia Prevalence):

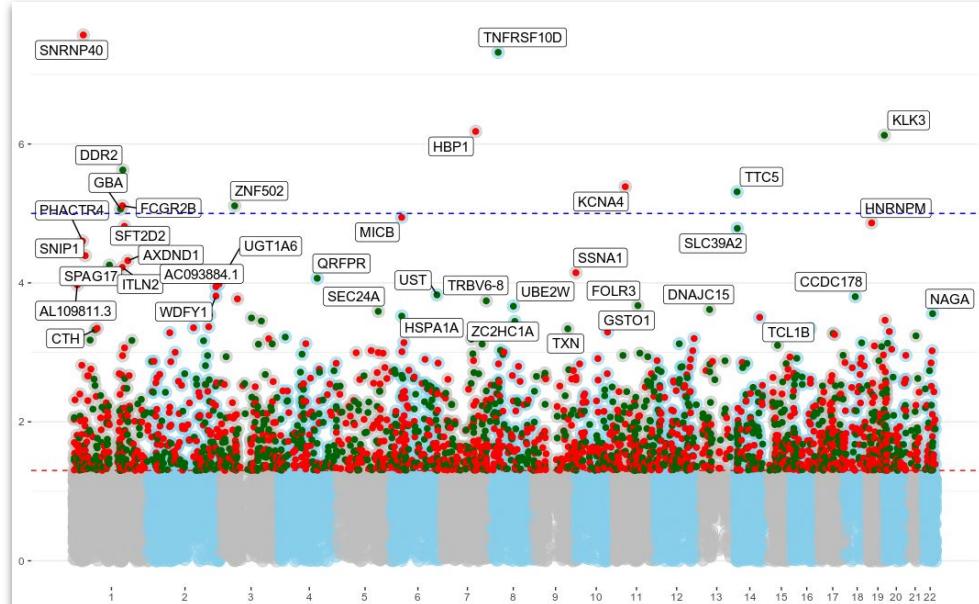
- a) Genes with **half of the species with certain dN/dS values** (value of “0”, ~10 species) were not included in the correlation analysis (a value of “0” biases the analysis because it tends to assign the strength of the correlations strength to particular species rather than to phylogenetic trends)
- b) Here we work with the removal of standardized residuals for the **raw values of the trait** (with several “0” observations, a log10-transformation before the correlation estimation does not lead to a roughly normal or uniform distribution).

Neoplasia Prevalence parametric p-values



Bonferroni
threshold

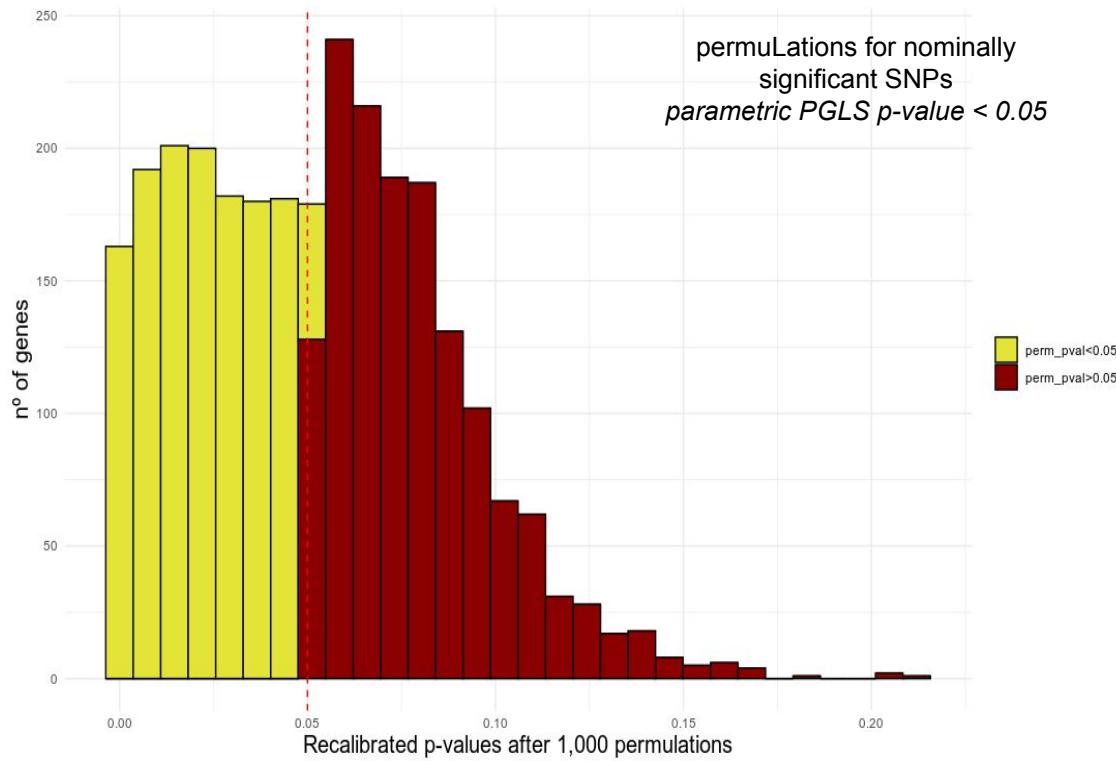
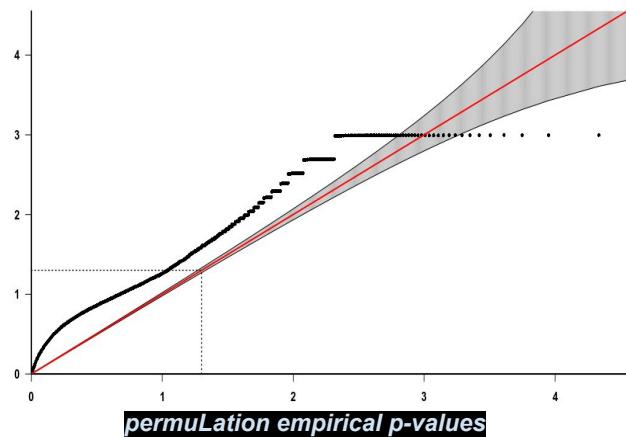
P-VAL <0.05
threshold



Some genes with roles as tumor suppressor and tumor necrosis factor as top hits

- + gene-trait correlations
- gene-trait correlations

Neoplasia Prevalence permuLated regressions



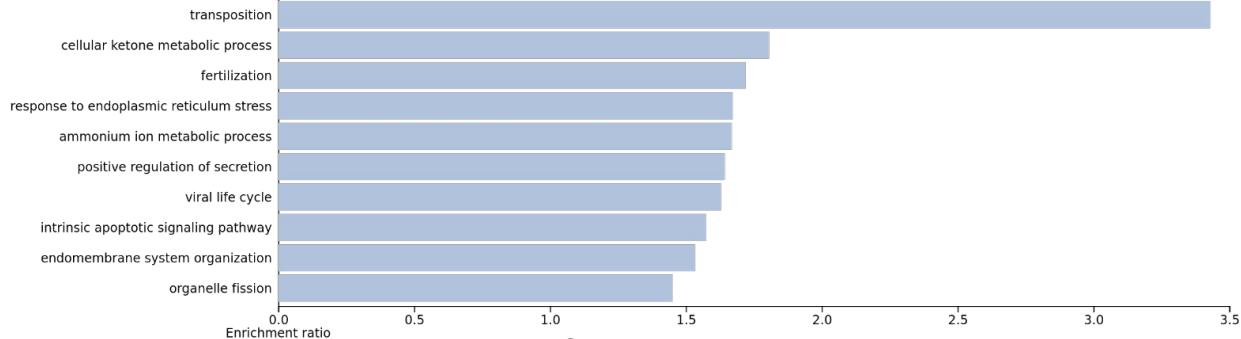
1,350 genes with **permuation p-value <0.05 & parametric PGLS p-value <0.05** (48.31% all nominally significant p-values)

Neoplasia Prevalence

functional enrichment

FDR ≤ 0.05 FDR > 0.05

Biological process



In-silico functional analysis

GO enrichment analysis

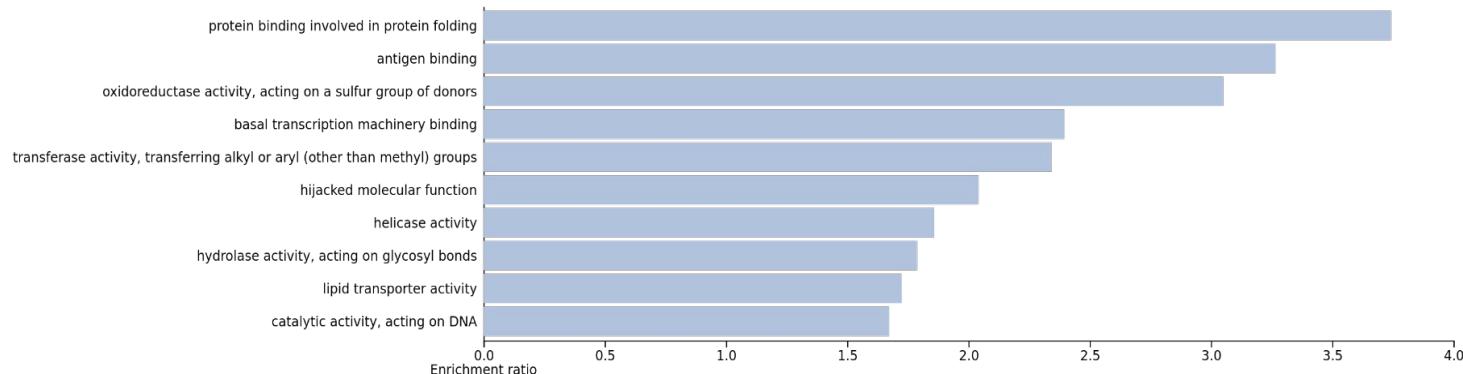
WebGestalt

<http://www.webgestalt.org/>

(Analysis performed upon Discovered CAAS)

Reference set to 16,343 genes

Molecular function



Neoplasia Prevalence

functional enrichment

FDR ≤ 0.05 FDR > 0.05

In-silico functional analysis

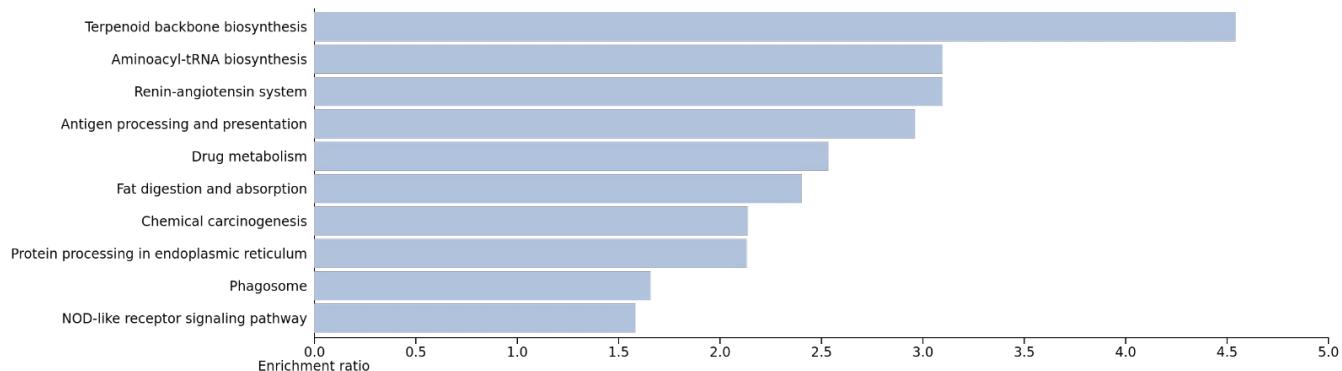
Pathway enrichment analysis

WebGestalt

<http://www.webgestalt.org/>

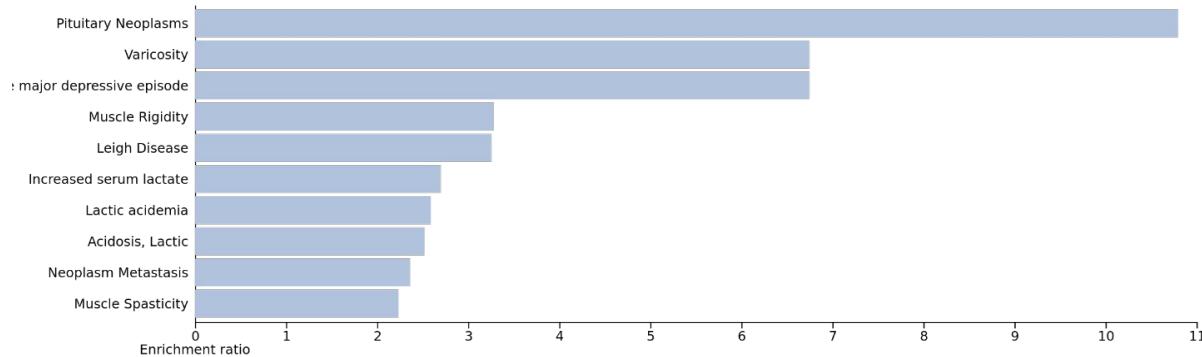
(Analysis performed upon Discovered CAAS)
Reference set to 16,343 genes

KEGG 2021: Biological pathways

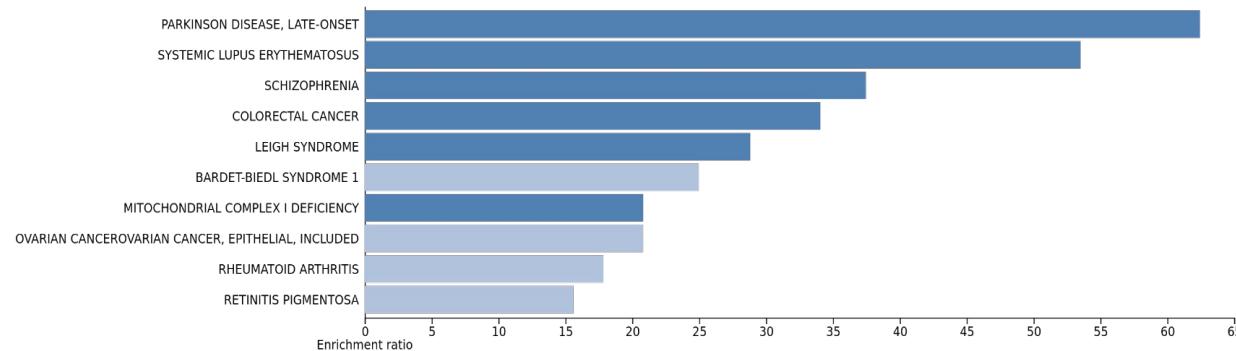


Neoplasia Prevalence functional enrichment

FDR ≤ 0.05 FDR > 0.05

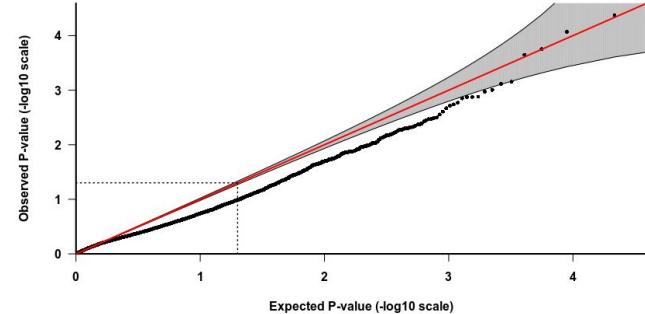
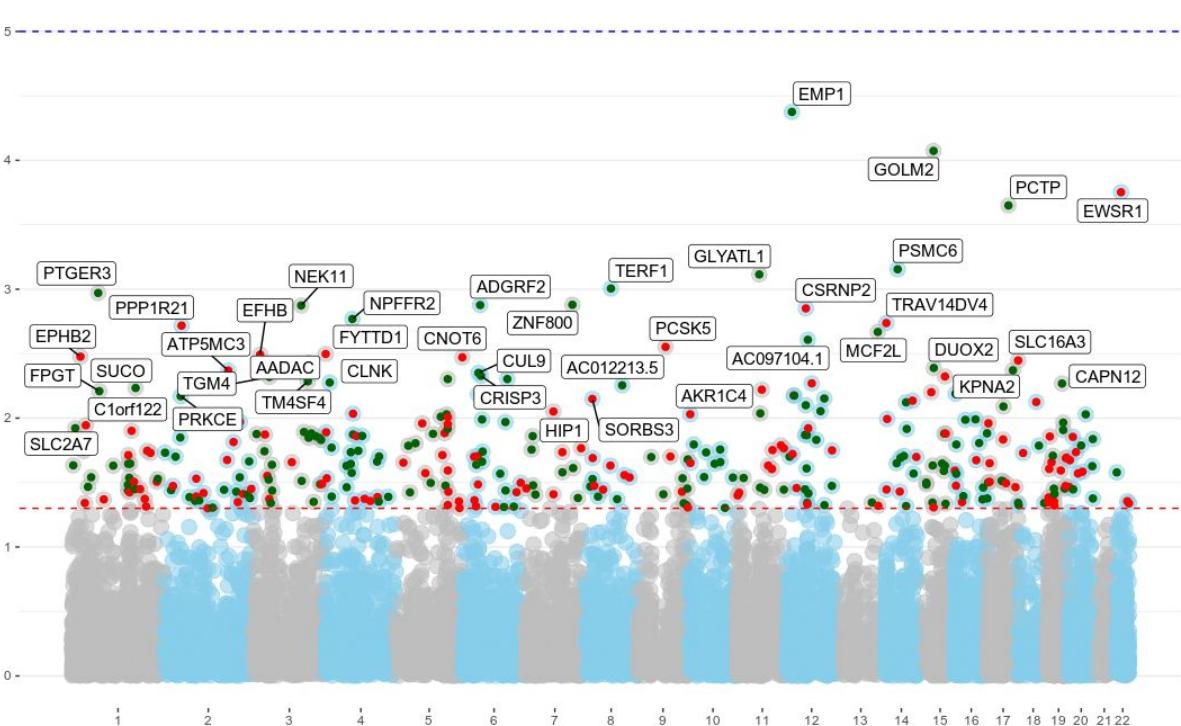


Disgenet & OMIM diseases



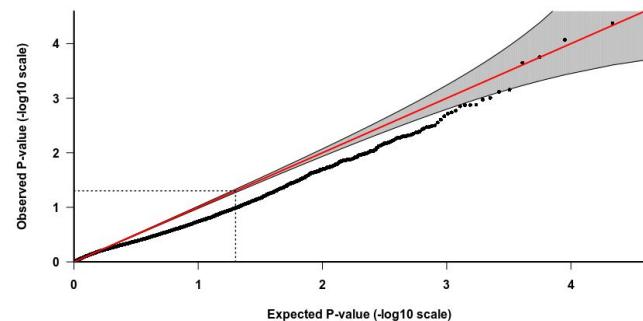
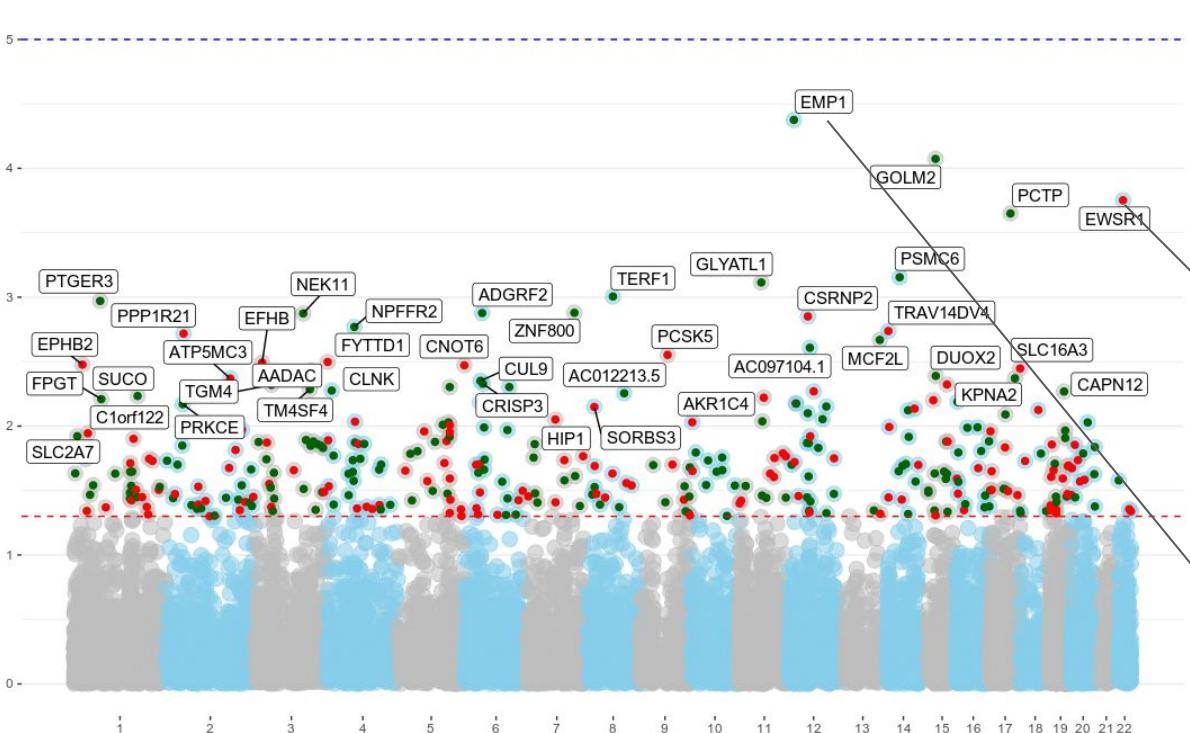
Pathway and disease enrichment analysis
WebGestalt
<http://www.webgestalt.org/>
(Analysis performed upon Discovered CAAS)
Reference set to 16,343 genes

Malignancy Prevalence parametric p-values



No excess of low p-values **but**
Top ranked genes with close
cancer links

Malignancy Prevalence parametric p-values



Reconstruction of Ewing Sarcoma Developmental Context from Mass-Scale Transcriptomics Reveals Characteristics of EWSR1-FLI1 Permissibility

by Henry E. Miller^{1,2} Aparna Gorhti^{1,2} Nicklas Bassani², Liesl A. Lawrence^{1,2}, Brian S. Iskra^{1,2} and Alexander J. R. Bishop^{1,2*}

¹ Department of Cell Systems and Anatomy, University of Texas Health at San Antonio, San Antonio, TX 78229, USA

² Greehey Children's Cancer Research Institute, University of Texas Health at San Antonio, San Antonio, TX 78229, USA

* Author to whom correspondence should be addressed.

Oncogene

Explore content ▾ About the journal ▾ Publish with us ▾

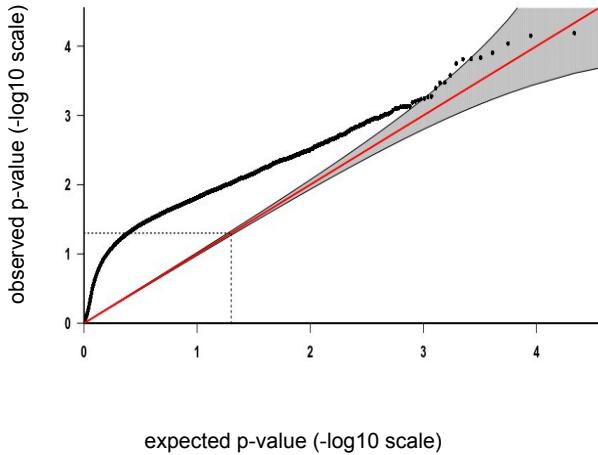
nature > oncogene > articles > article

Article | Open Access | Published: 04 June 2018

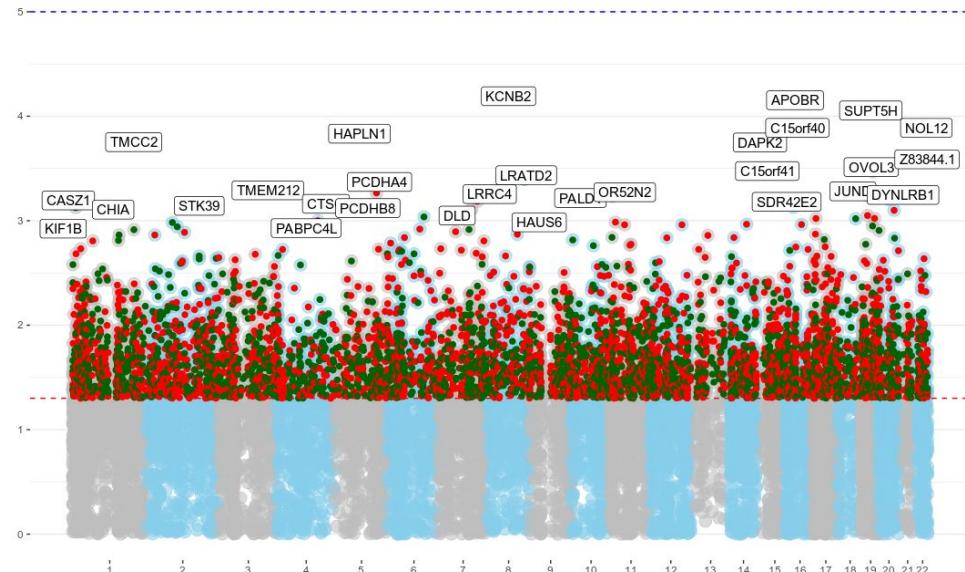
Epithelial membrane protein 1 promotes tumor metastasis by enhancing cell migration via copine-III and Rac1

Benignancy Prevalence parametric p-values

Bonferroni
threshold

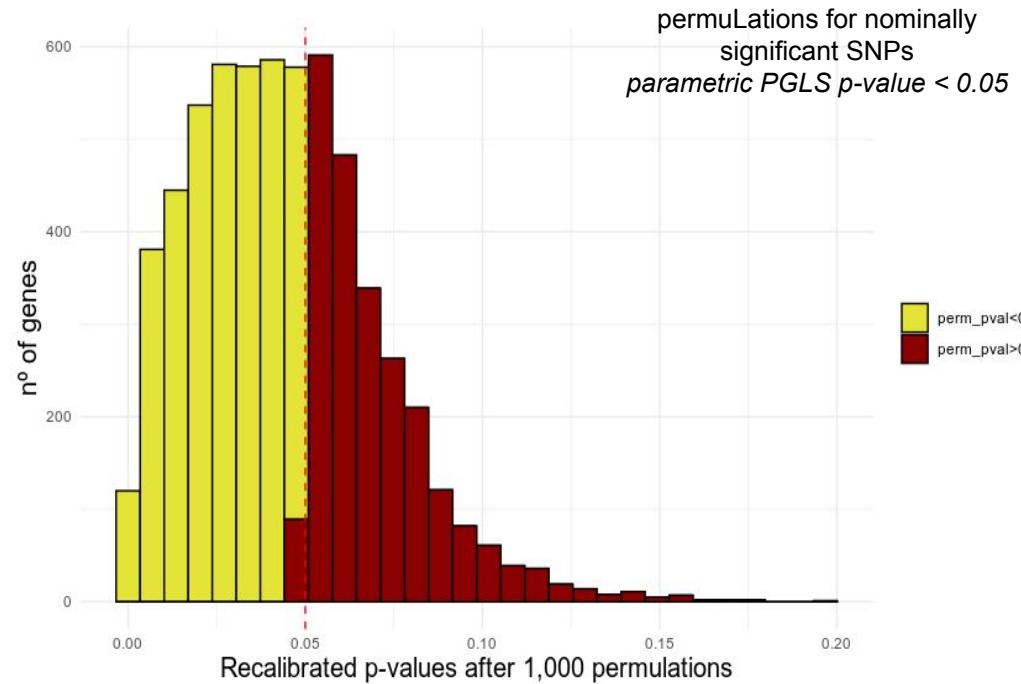
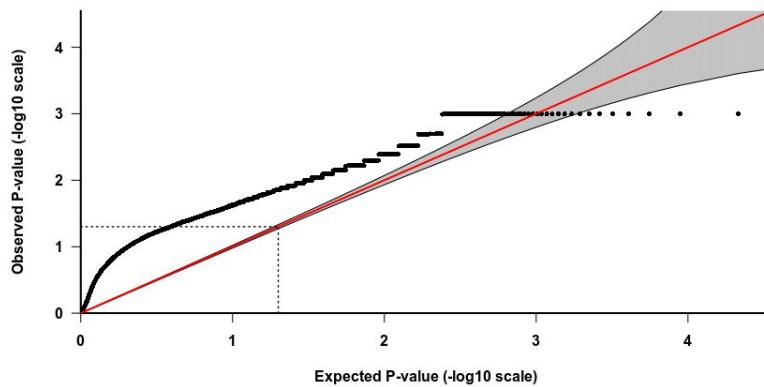


P-VAL <0.05
threshold



- + gene-trait correlations
- gene-trait correlations

Benignancy Prevalence permulated associations



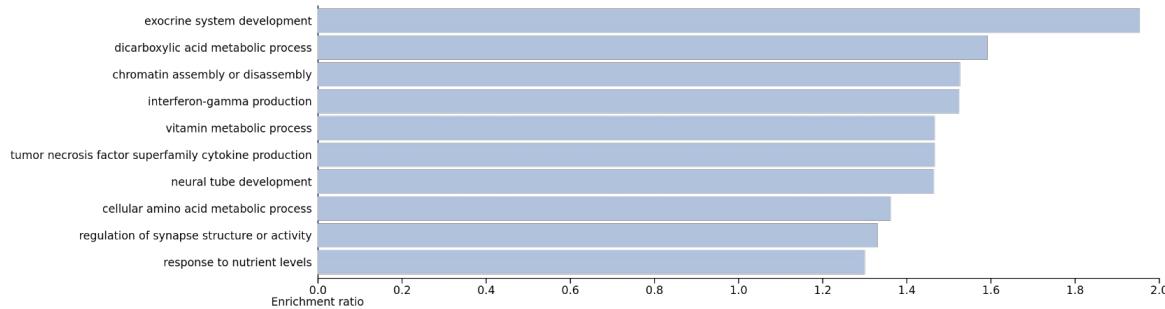
3,718 genes with **permulation p-value <0.05** & **parametric PGLS p-value <0.05** (60.6% all nominally significant p-values)

Benignancy Prevalence

functional enrichment

FDR ≤ 0.05 FDR > 0.05

Biological process



In-silico functional analysis

GO enrichment analysis

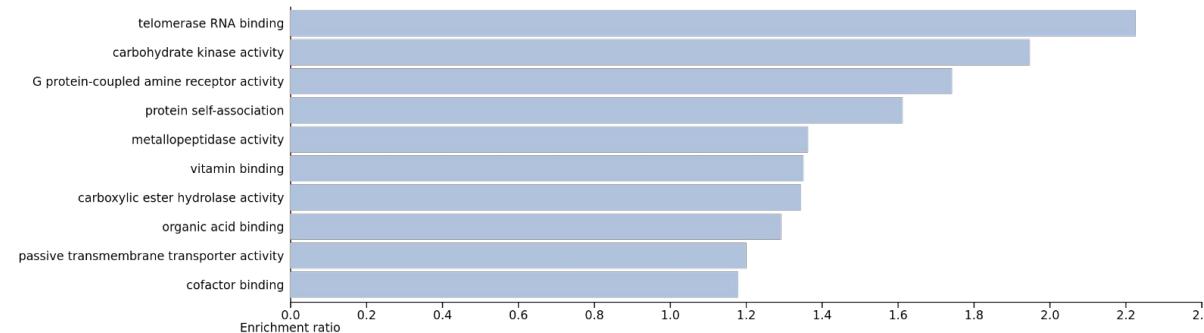
WebGestalt

<http://www.webgestalt.org/>

(Analysis performed upon Discovered CAAS)

Reference set to 16,343 genes

Molecular function

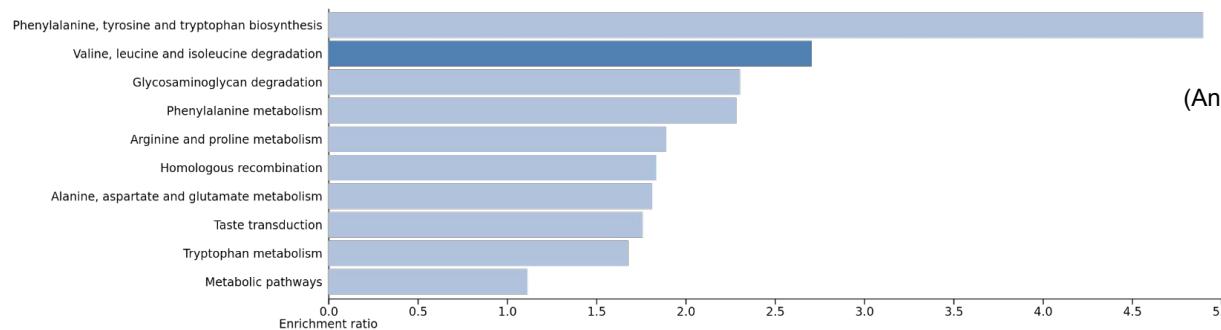


Benignancy Prevalence

functional enrichment

FDR ≤ 0.05 FDR > 0.05

KEGG 2021: Biological pathways



In-silico functional analysis

Pathway and disease analysis

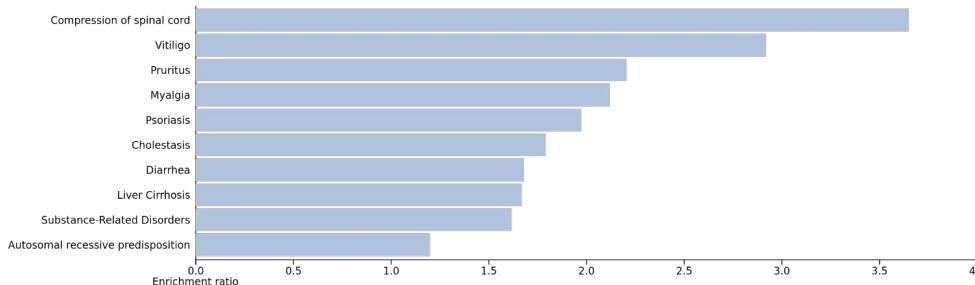
WebGestalt

<http://www.webgestalt.org/>

(Analysis performed upon Discovered CAAS)

Reference set to 16,343 genes

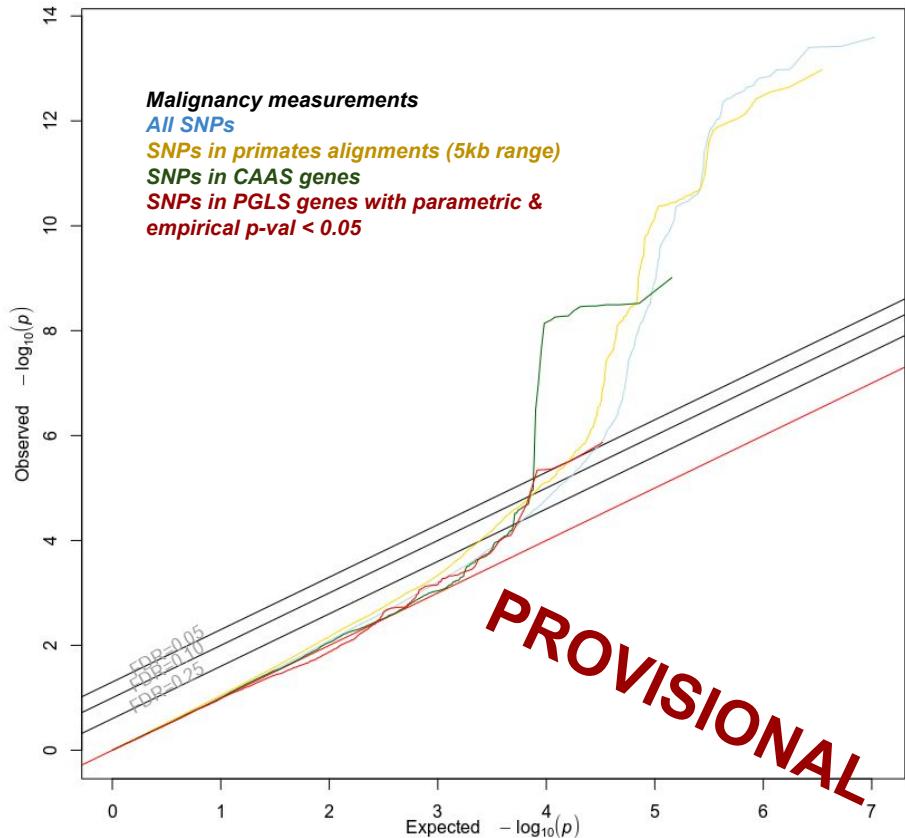
Disgenet diseases



Future steps

- Look for more **neoplasm-related summary statistics from human GWAS** and ascertain the proportions of heritability explained by our validated genesets.

Example:
Cancers diagnosed by doctor:
Malignant, primary site
(GWAS Atlas ID: 3329)



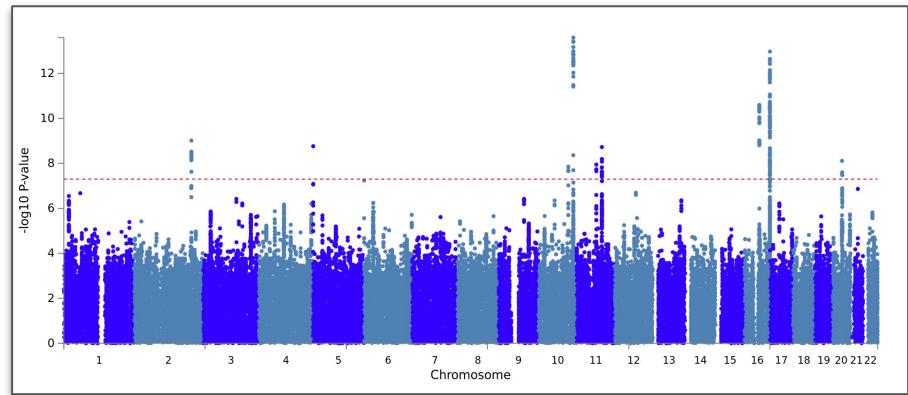
Future steps

- Look for more **neoplasm-related summary statistics from human GWAS** and ascertain the proportions of heritability explained by our validated genesets.

Example:

Cancers diagnosed by doctor:

Malignant, primary site
(GWAS Atlas ID: 3329)



PROVISIONAL

Category	Prop SNPs	Prop h2	Enrichment	Enrichment_p
PGLS p < 0.05	0.0738	0.1482	2.009	0.0898

Final Summary

- ***Neoplasia Prevalence*** appears is the trait with the largest # of links with neoplasm annotation databases fort both, the CAAS and PGLS methodologies
- Top ranked genes from *Neoplasia and Malignancy Prevalence* correlation analysis also play roles in some cancer subtypes (e.g., *KLK3 / TNFRFS10D / HBP1 /EWSR1..*)
- Most of the associated genesets are enriched for *aminoacid synthesis, fatty acid absoption, inflamassome complex and apoptosis*-related pathways biological categories

Acknowledgements

Comparative and Evolutionary Genomics Lab



Gerard Muntané



Arcadi Navarro



David Juan



Thanks!!

Any
questions?



