Navigating the Complexities of the Human Oncoproteome with the SigNET KnowledgeBank

Presented by Steven Pelech, Ph.D.

Professor, Department of Medicine,
University of British Columbia
President & CSO, Kinexus Bioinformatics Corp.

Presented at the 2014 BioConference Live Clinical Diagnostics and Research Conference, 2014 May 29

U.S. Human Cancers

Site	2012 Occurrence		2006-2008 Risk	
	Male	Female	Male	Female
Prostate	29%		1 in 6	
Breast		29%		1 in 8
Lung & bronchus	14%	14%	1 in 13	1 in 16
Colon & rectum	9%	9%	1 in 19	1 in 20
Urinary bladder‡	7%	2%	1 in 26	1 in 87
Melanoma	5%	4%	1 in 36	
Uterine corpus		6%		1 in 38
Thyroid	1%	5%		
Non-Hodgkin lymphoma	4%	4%	1 in 43	1 in 51
Kidney & renal pelvis	5%	5%	1 in 51	
Leukemia	3%	3%	1 in 64	
Oral cavity	3%	2%	1 in 69	
Pancreas	3%	3%		1 in 69
Ovary		3%		1 in 71
All other sites	17%	11%		
All sites	100%	100%	1 in 2	1 in 3

U.S. Human Cancers

Relative Survival* (%) During Five Periods by Cancer Site

Site	1975-1977	1983-1985	1987-1989	1992-1998	2001-2007
Breast (female)	75	78	84	86	90
Colon	51	57	60	62	65
Leukemia	34	41	43	46	57
Lung & bronchus	12	14	13	15	16
Melanoma	82	85	88	89	93
Non-Hodgkin lymphoma	47	54	51	55	70
Ovary	36	41	38	53	44
Pancreas	2	3	4	4	6
Prostate	68	75	83	97	100
Rectum	48	57	58	62	68
Urinary bladder	73	78	79	82	80
All sites	50	52	56	62	67

Results After more than 50 Years of Intense NIH Research Funding

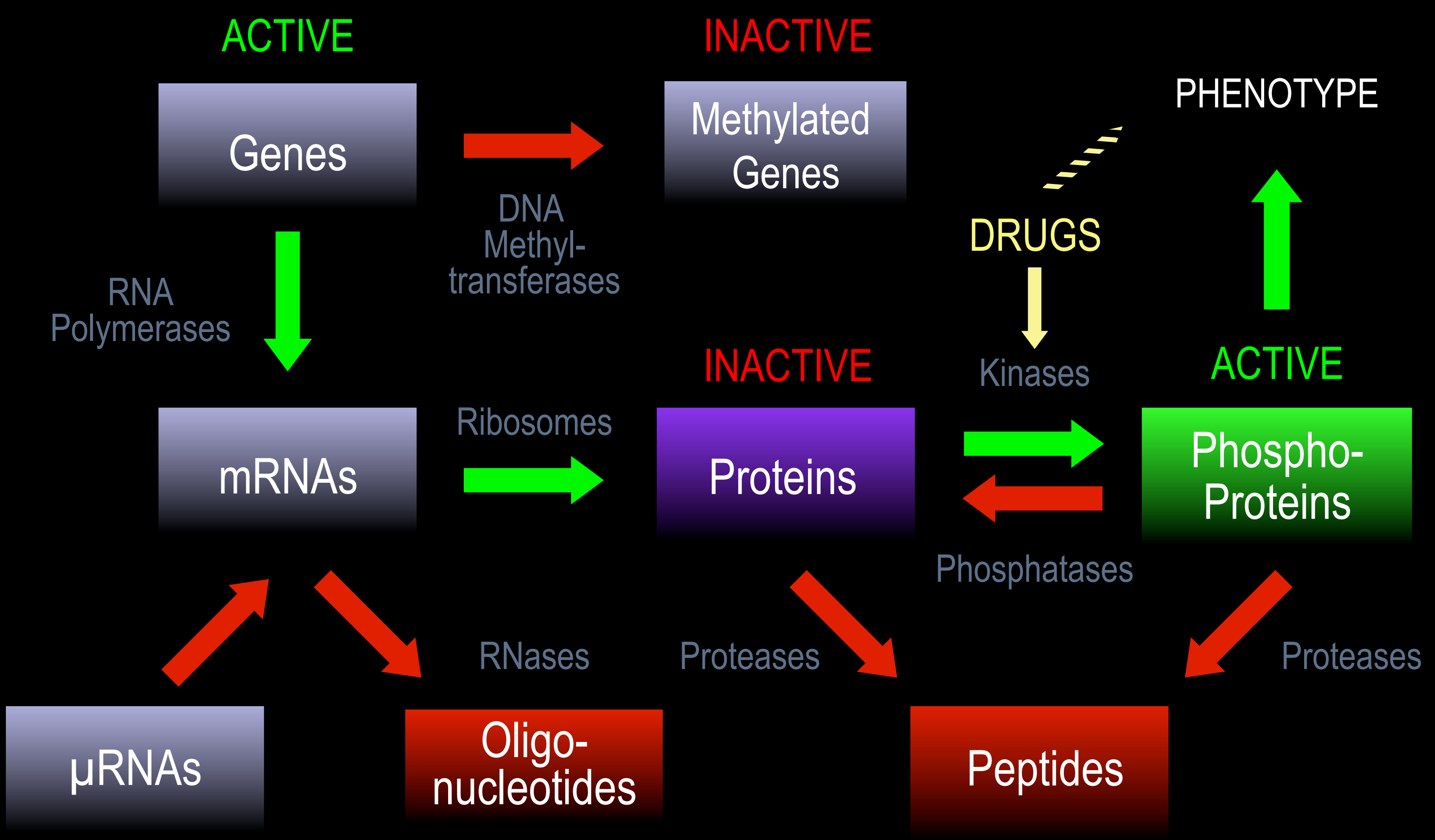
- 5 year survival rates have only increased by less than 20%
- About of third of cancer patients still die within 5 years
- The lion share of cancer research funding actually goes to the most treatable forms of cancer
- The human genome has been sequenced for more than a decade and we still do not have much knowledge about what over 40% of the estimated 21,500 protein-coding genes actually do
- Over 3000 genes have been linked to human cancers
- Every human primary tumour may be unique in its etiology and heterogeneous with respect to its mutations

Presentation Objectives

- The complexity of the human genome and proteome is a major barrier for developing improved diagnoses and treatments of cancer towards the aim of personalized medicine

- This presentation will:
  - Summarize the efforts of Kinexus to create open-access bioinformatics databases and websites for the study of cell signalling systems under normal and neoplastic conditions
  - Summarize some of the results of queries of these databases with respect to cancer gene mutation and expression in diverse human cancers
  - Introduce the concept of “Tumour Requiring Proteins”
  - Highlight the importance of protein kinases as cancer diagnostic markers and therapeutic drug targets

Phenotype Reflects Functional Protein



ACTIVE

Genes

INACTIVE

Methylated Genes

DNA Methyltransferases

PHENOTYPE

DRUGS

RNA Polymerases

mRNAs

Ribosomes

INACTIVE

Proteins

Kinases

ACTIVE

Phospho-Proteins

Phosphatases

RNases

Proteases

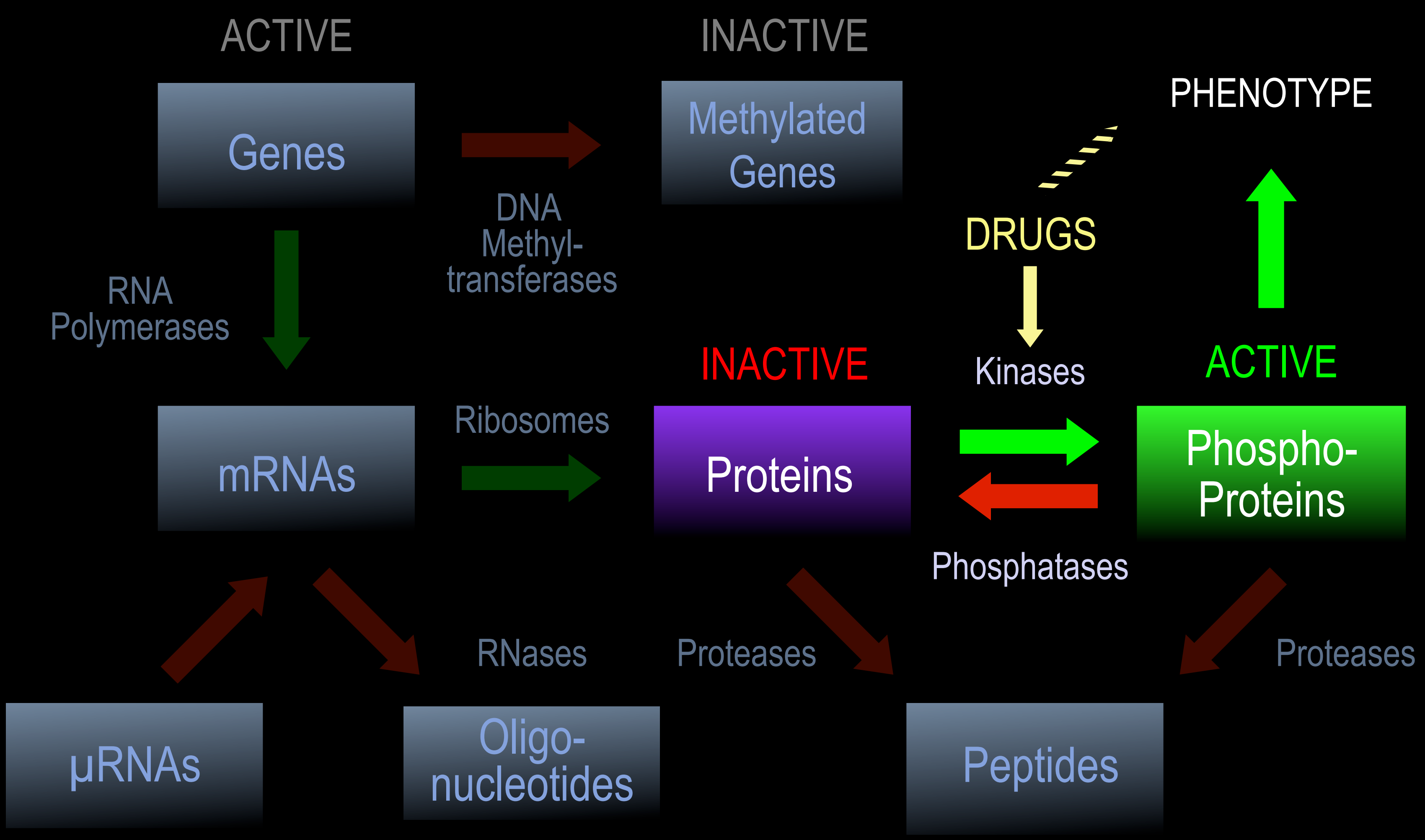
Proteases

µRNAs

Oligonucleotides

Peptides

Phenotype Reflects Functional Protein



Human Protein Phosphorylation

PROTEOME

~21,500 total protein genes

KINEOME

>536 protein kinase genes

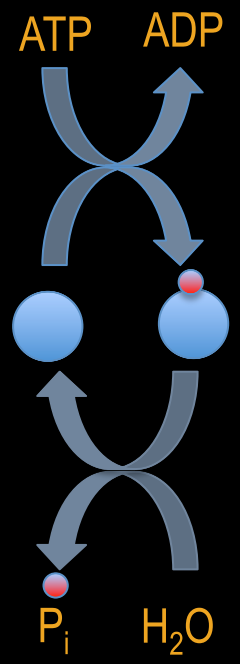
>1000 catalytically active protein kinase isoforms

PHOSPHO-PROTEOME

~1,000,000 predicted phosphorylation sites (>200,000 experimentally confirmed)

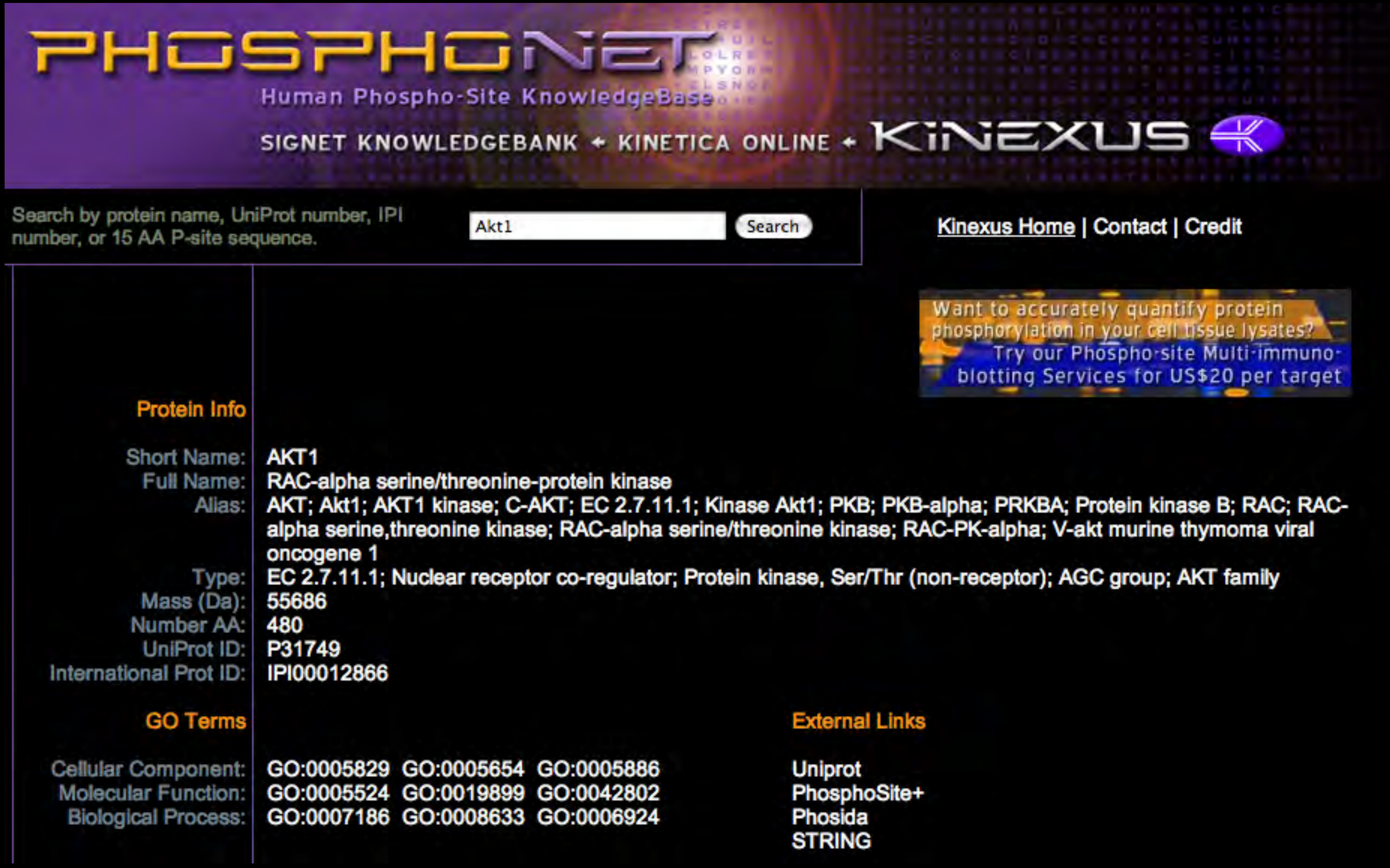
PHOSPHATOME

>156 protein phosphatase genes



Phosphosite Identification in Proteins

www.phosphonet.ca



Search by protein name, UniProt number, IPI number, or 15 AA P-site sequence.

Protein Info

Short Name:	AKT1
Full Name:	RAC-alpha serine/threonine-protein kinase
Alias:	AKT; Akt1; AKT1 kinase; C-AKT; EC 2.7.11.1; Kinase Akt1; PKB; PKB-alpha; PRKBA; Protein kinase B; RAC; RAC-alpha serine,threonine kinase; RAC-alpha serine/threonine kinase; RAC-PK-alpha; V-akt murine thymoma viral oncogene 1
Type:	EC 2.7.11.1; Nuclear receptor co-regulator; Protein kinase, Ser/Thr (non-receptor); AGC group; AKT family
Mass (Da):	55686
Number AA:	480
UniProt ID:	P31749
International Prot ID:	IPI00012866

GO Terms

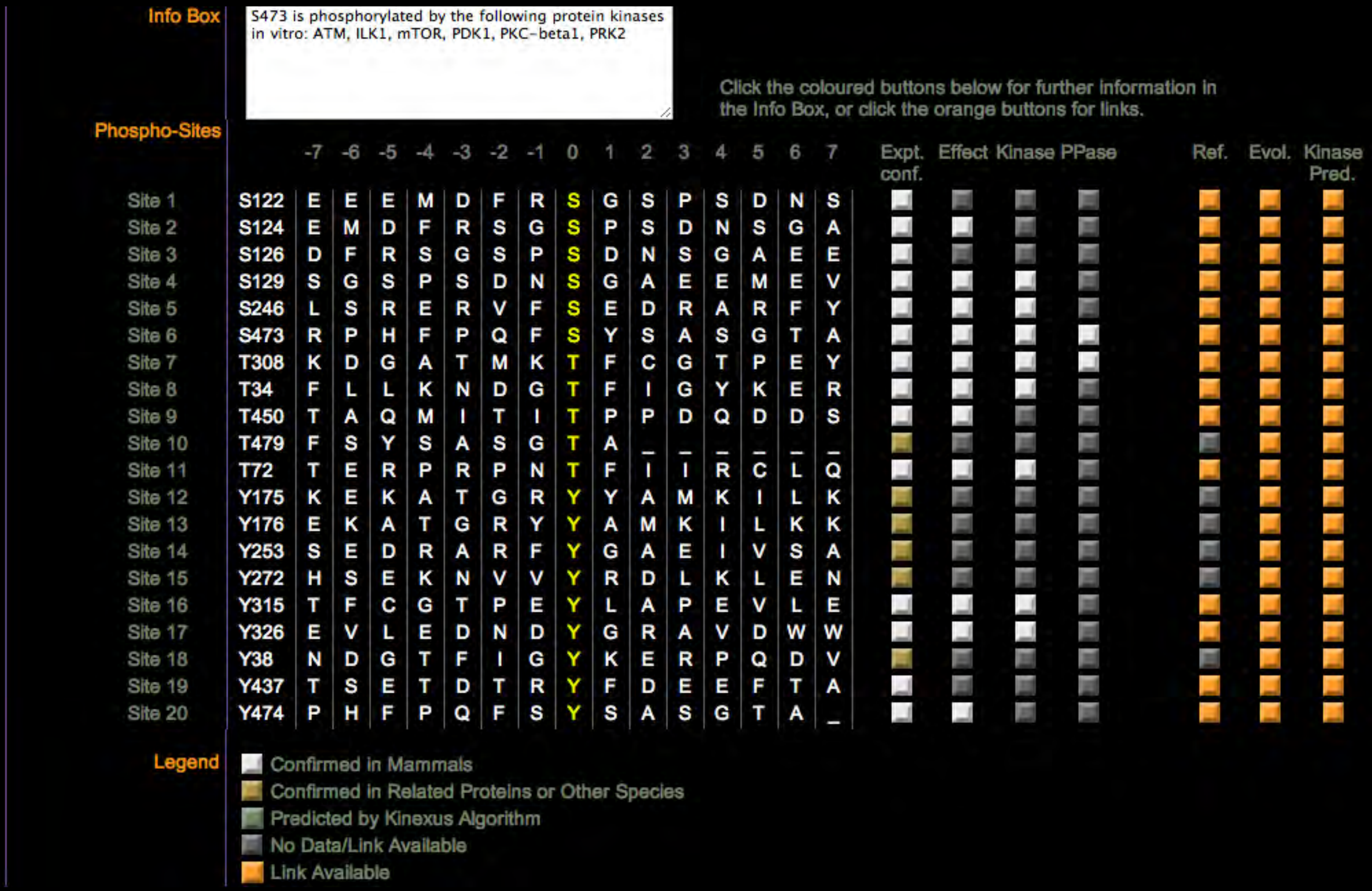
Cellular Component:	GO:0005829	GO:0005654	GO:0005886
Molecular Function:	GO:0005524	GO:0019899	GO:0042802
Biological Process:	GO:0007186	GO:0008633	GO:0006924

External Links

Uniprot
PhosphoSite+
Phosida
STRING

Phosphosite Identification in Proteins

www.phosphonet.ca



Info Box

S473 is phosphorylated by the following protein kinases in vitro: ATM, ILK1, mTOR, PDK1, PKC-beta1, PRK2

Click the coloured buttons below for further information in the Info Box, or click the orange buttons for links.

Phospho-Sites

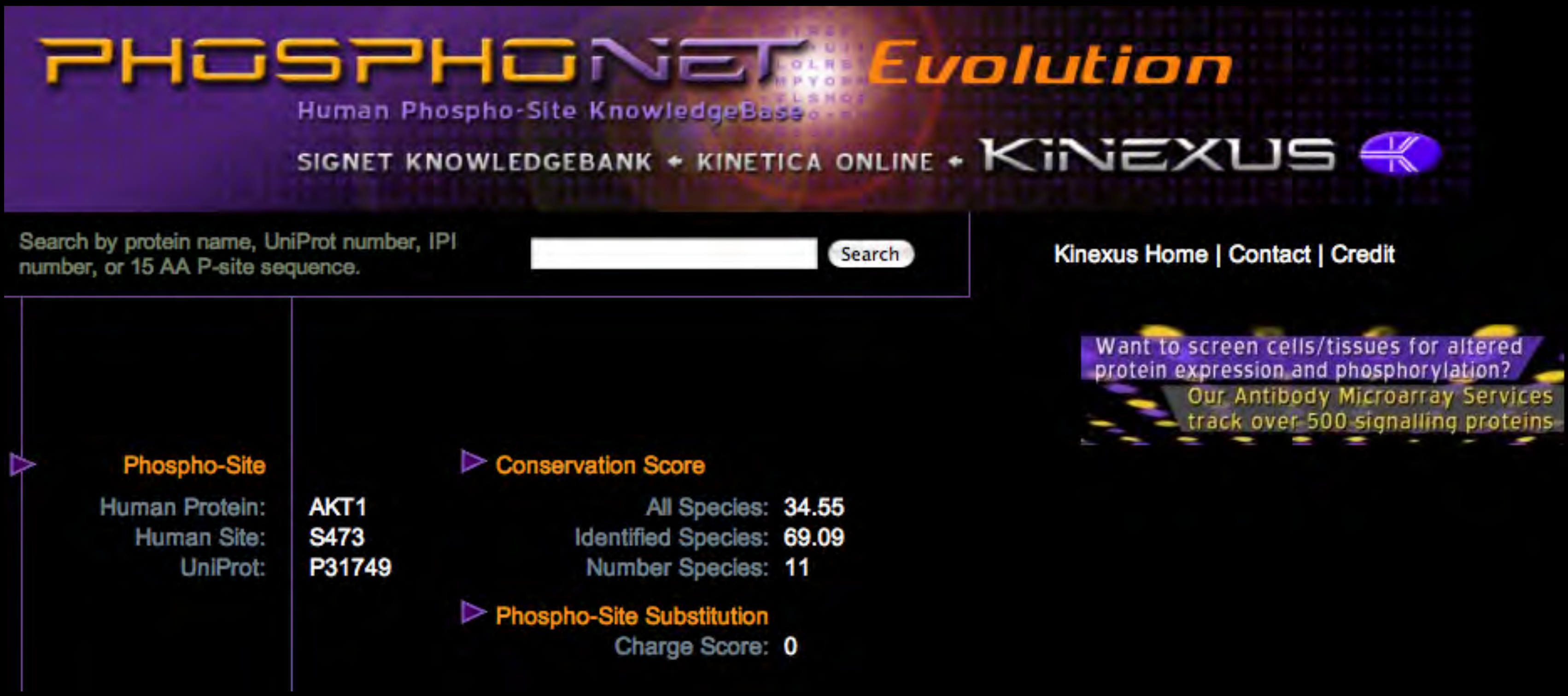
		-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	Expt. conf.	Effect	Kinase	PPase	Ref.	Evol.	Kinase Pred.
Site 1	S122	E	E	E	M	D	F	R	S	G	S	P	S	D	N	S	⬜	⬛	⬛	⬛	🟧	🟧	🟧
Site 2	S124	E	M	D	F	R	S	G	S	P	S	D	N	S	G	A	⬜	⬜	⬛	⬛	🟧	🟧	🟧
Site 3	S126	D	F	R	S	G	S	P	S	D	N	S	G	A	E	E	⬜	⬛	⬛	⬛	🟧	🟧	🟧
Site 4	S129	S	G	S	P	S	D	N	S	G	A	E	E	M	E	V	⬜	⬜	⬜	⬛	🟧	🟧	🟧
Site 5	S246	L	S	R	E	R	V	F	S	E	D	R	A	R	F	Y	⬜	⬜	⬜	⬛	🟧	🟧	🟧
Site 6	S473	R	P	H	F	P	Q	F	S	Y	S	A	S	G	T	A	⬜	⬜	⬜	⬜	🟧	🟧	🟧
Site 7	T308	K	D	G	A	T	M	K	T	F	C	G	T	P	E	Y	⬜	⬜	⬜	⬜	🟧	🟧	🟧
Site 8	T34	F	L	L	K	N	D	G	T	F	I	G	Y	K	E	R	⬜	⬜	⬜	⬛	🟧	🟧	🟧
Site 9	T450	T	A	Q	M	I	T	I	T	P	P	D	Q	D	D	S	⬜	⬜	⬛	⬛	🟧	🟧	🟧
Site 10	T479	F	S	Y	S	A	S	G	T	A	_	_	_	_	_	_	🟨	⬛	⬛	⬛	⬛	🟧	🟧
Site 11	T72	T	E	R	P	R	P	N	T	F	I	I	R	C	L	Q	⬜	⬜	⬜	⬛	🟧	🟧	🟧
Site 12	Y175	K	E	K	A	T	G	R	Y	Y	A	M	K	I	L	K	🟨	⬛	⬛	⬛	⬛	🟧	🟧
Site 13	Y176	E	K	A	T	G	R	Y	Y	A	M	K	I	L	K	K	🟨	⬛	⬛	⬛	⬛	🟧	🟧
Site 14	Y253	S	E	D	R	A	R	F	Y	G	A	E	I	V	S	A	🟨	⬛	⬛	⬛	⬛	🟧	🟧
Site 15	Y272	H	S	E	K	N	V	V	Y	R	D	L	K	L	E	N	🟨	⬛	⬛	⬛	⬛	🟧	🟧
Site 16	Y315	T	F	C	G	T	P	E	Y	L	A	P	E	V	L	E	⬜	⬜	⬜	⬛	🟧	🟧	🟧
Site 17	Y326	E	V	L	E	D	N	D	Y	G	R	A	V	D	W	W	⬜	⬜	⬜	⬛	🟧	🟧	🟧
Site 18	Y38	N	D	G	T	F	I	G	Y	K	E	R	P	Q	D	V	🟨	⬛	⬛	⬛	⬛	🟧	🟧
Site 19	Y437	T	S	E	T	D	T	R	Y	F	D	E	E	F	T	A	⬜	⬛	⬛	⬛	🟧	🟧	🟧
Site 20	Y474	P	H	F	P	Q	F	S	Y	S	A	S	G	T	A	_	⬜	⬜	⬛	⬛	🟧	🟧	🟧

Legend

- ⬜ Confirmed in Mammals
- 🟨 Confirmed in Related Proteins or Other Species
- 🟩 Predicted by Kinexus Algorithm
- ⬛ No Data/Link Available
- 🟧 Link Available

Evolutionary Analysis of Phosphosites

www.phosphonet.ca



Search by protein name, UniProt number, IPI number, or 15 AA P-site sequence.

Want to screen cells/tissues for altered protein expression and phosphorylation? Our Antibody Microarray Services track over 500 signalling proteins

Phospho-Site

Human Protein: AKT1
Human Site: S473
UniProt: P31749

Conservation Score

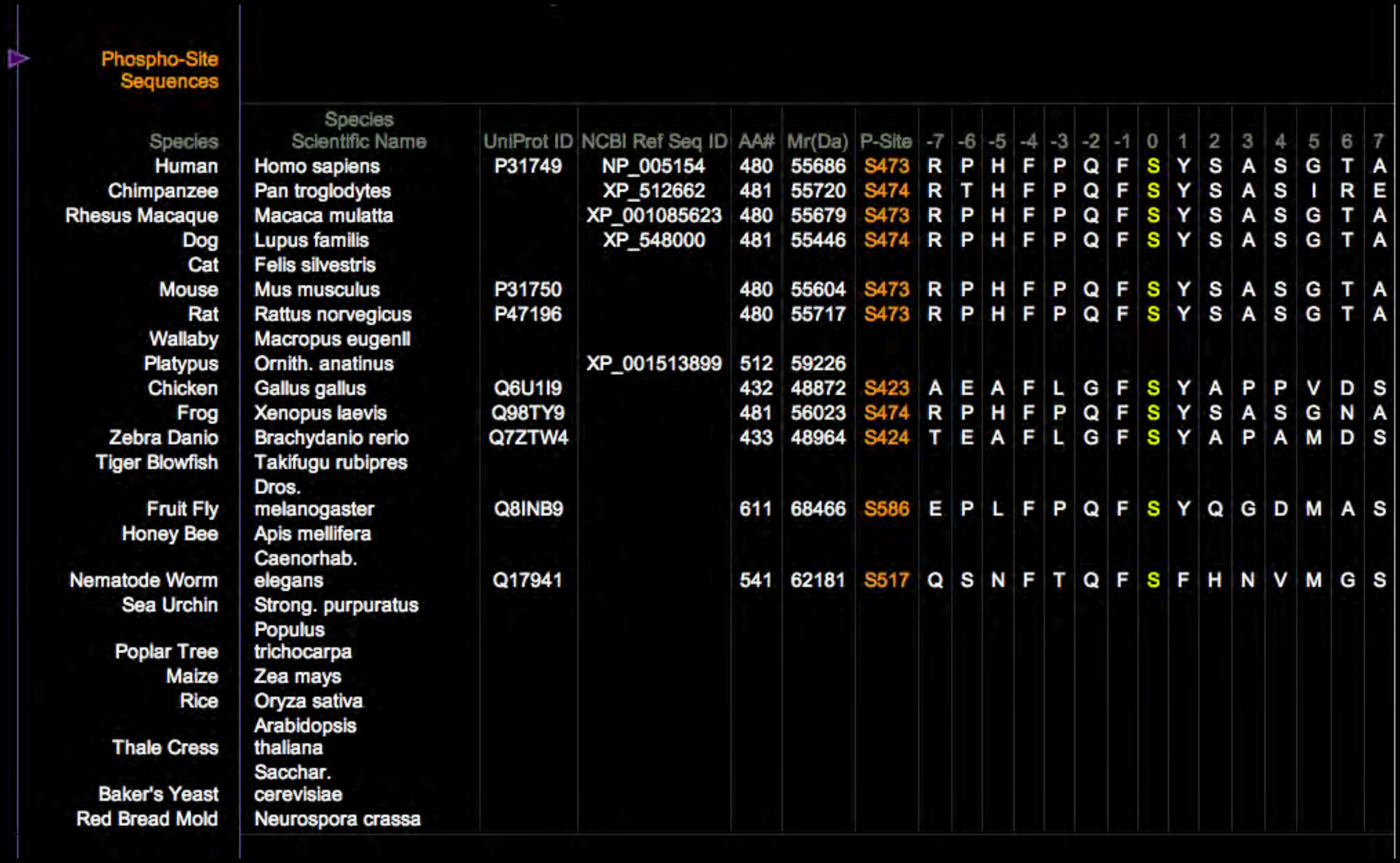
All Species: 34.55
Identified Species: 69.09
Number Species: 11

Phospho-Site Substitution

Charge Score: 0

Evolutionary Analysis of Phosphosites

www.phosphonet.ca



Phospho-Site Sequences

Species	Species Scientific Name	UniProt ID	NCBI Ref Seq ID	AA#	Mr(Da)	P-Site	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
Human	Homo sapiens	P31749	NP_005154	480	55686	S473	R	P	H	F	P	Q	F	S	Y	S	A	S	G	T	A
Chimpanzee	Pan troglodytes		XP_512662	481	55720	S474	R	T	H	F	P	Q	F	S	Y	S	A	S	I	R	E
Rhesus Macaque	Macaca mulatta		XP_001085623	480	55679	S473	R	P	H	F	P	Q	F	S	Y	S	A	S	G	T	A
Dog	Lupus familis		XP_548000	481	55446	S474	R	P	H	F	P	Q	F	S	Y	S	A	S	G	T	A
Cat	Felis silvestris																				
Mouse	Mus musculus	P31750		480	55604	S473	R	P	H	F	P	Q	F	S	Y	S	A	S	G	T	A
Rat	Rattus norvegicus	P47196		480	55717	S473	R	P	H	F	P	Q	F	S	Y	S	A	S	G	T	A
Wallaby	Macropus eugenll																				
Platypus	Ornith. anatinus		XP_001513899	512	59226																
Chicken	Gallus gallus	Q6U1I9		432	48872	S423	A	E	A	F	L	G	F	S	Y	A	P	P	V	D	S
Frog	Xenopus laevis	Q98TY9		481	56023	S474	R	P	H	F	P	Q	F	S	Y	S	A	S	G	N	A
Zebra Danio	Brachydanio rerio	Q7ZTW4		433	48964	S424	T	E	A	F	L	G	F	S	Y	A	P	A	M	D	S
Tiger Blowfish	Takifugu rubipres																				
Fruit Fly	Dros. melanogaster	Q8INB9		611	68466	S586	E	P	L	F	P	Q	F	S	Y	Q	G	D	M	A	S
Honey Bee	Apis mellifera																				
Nematode Worm	Caenorhab. elegans	Q17941		541	62181	S517	Q	S	N	F	T	Q	F	S	F	H	N	V	M	G	S
Sea Urchin	Strong. purpuratus																				
Poplar Tree	Populus trichocarpa																				
Maize	Zea mays																				
Rice	Oryza sativa																				
Thale Cress	Arabidopsis thaliana																				
Baker's Yeast	Sacchar. cerevisiae																				
Red Bread Mold	Neurospora crassa																				

Evolutionary Analysis of Phosphosites

www.phosphonet.ca

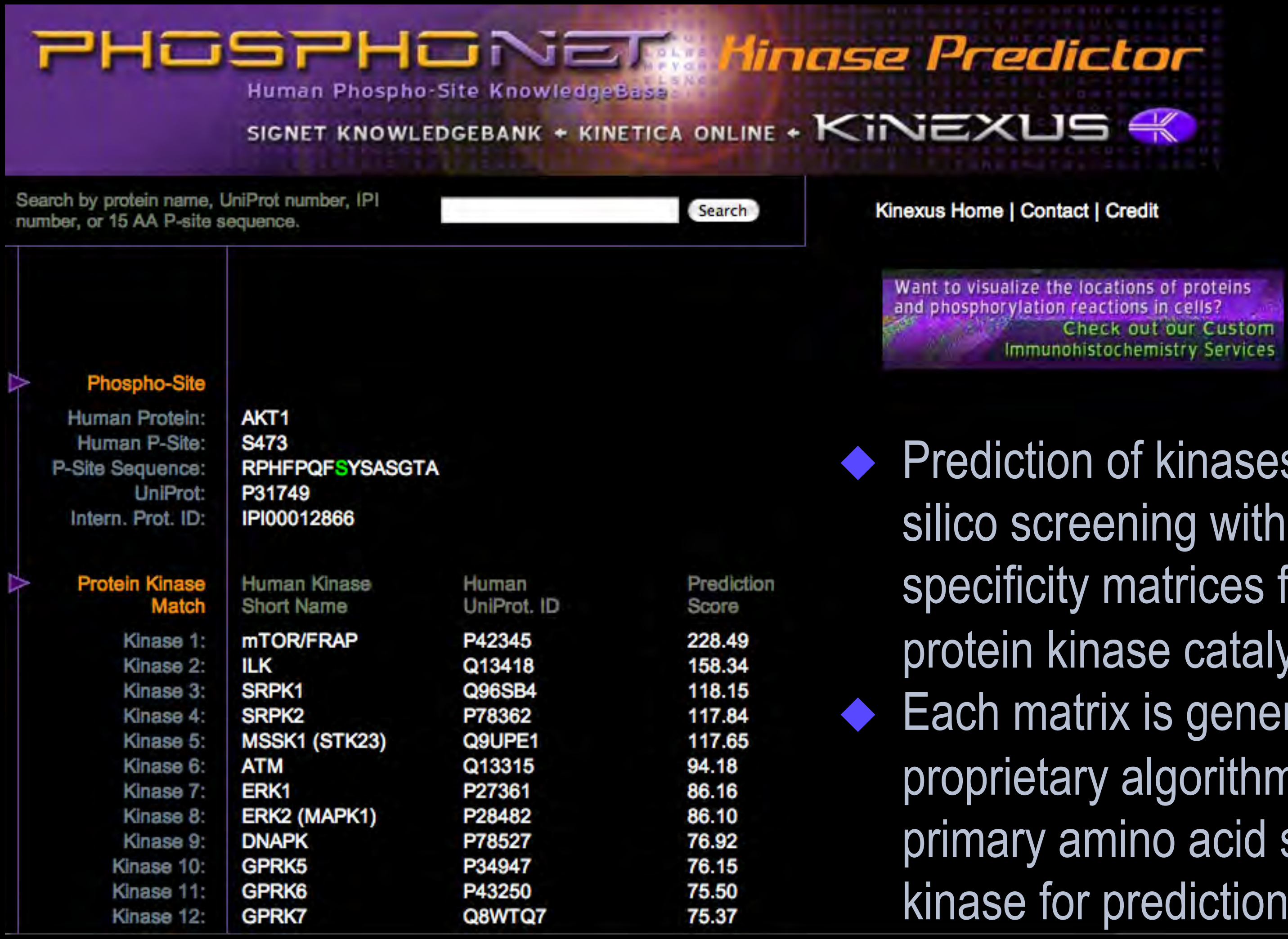
Conservation

Percent	Human	Chimpanzee	Rhesus Macaque	Dog	Cat	Mouse	Rat	Wallaby	Platypus	Chicken	Frog	Zebra Danio	Tiger Blowfish	Fruit Fly	Honey Bee	Nematode Worm	Sea Urchin
Protein Identity:	100	81.5	99.5	93.9	N.A.	98.1	98.1	N.A.	73	39.7	93.1	39.3	N.A.	49.5	N.A.	52.5	N.A.
Protein Similarity:	100	92.1	99.5	95.2	N.A.	98.7	99.1	N.A.	80.4	56.8	96.2	57.5	N.A.	61.7	N.A.	67.8	N.A.
P-Site Identity:	100	73.3	100	100	N.A.	100	100	N.A.	0	26.6	93.3	26.6	N.A.	46.6	N.A.	26.6	N.A.
P-Site Similarity:	100	73.3	100	100	N.A.	100	100	N.A.	0	40	93.3	46.6	N.A.	53.3	N.A.	53.3	N.A.

Percent	Poplar Tree	Maize	Rice	Thale Cress	Baker's Yeast	Red Bread Mold
Protein Identity:	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Protein Similarity:	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
P-Site Identity:	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
P-Site Similarity:	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

KINEXUS Kinase Substrate Phosphosite Prediction

www.phosphonet.ca



PHOSPHONET Kinase Predictor

Human Phospho-Site KnowledgeBase

SIGNET KNOWLEDGEBANK ◆ KINETICA ONLINE ◆ KINEXUS

Search by protein name, UniProt number, IPI number, or 15 AA P-site sequence.

Kinexus Home | Contact | Credit

Want to visualize the locations of proteins and phosphorylation reactions in cells? Check out our Custom Immunohistochemistry Services

Phospho-Site

Human Protein:	AKT1
Human P-Site:	S473
P-Site Sequence:	RPHFPQFSYSASGTA
UniProt:	P31749
Intern. Prot. ID:	IPI00012866

Protein Kinase Match	Human Kinase Short Name	Human UniProt. ID	Prediction Score
Kinase 1:	mTOR/FRAP	P42345	228.49
Kinase 2:	ILK	Q13418	158.34
Kinase 3:	SRPK1	Q96SB4	118.15
Kinase 4:	SRPK2	P78362	117.84
Kinase 5:	MSSK1 (STK23)	Q9UPE1	117.65
Kinase 6:	ATM	Q13315	94.18
Kinase 7:	ERK1	P27361	86.16
Kinase 8:	ERK2 (MAPK1)	P28482	86.10
Kinase 9:	DNAPK	P78527	76.92
Kinase 10:	GPRK5	P34947	76.15
Kinase 11:	GPRK6	P43250	75.50
Kinase 12:	GPRK7	Q8WTQ7	75.37

- Prediction of kinases is based in silico screening with phospho-site specificity matrices for 500 human protein kinase catalytic domains
- Each matrix is generated with a proprietary algorithm that uses the primary amino acid structure of the kinase for prediction

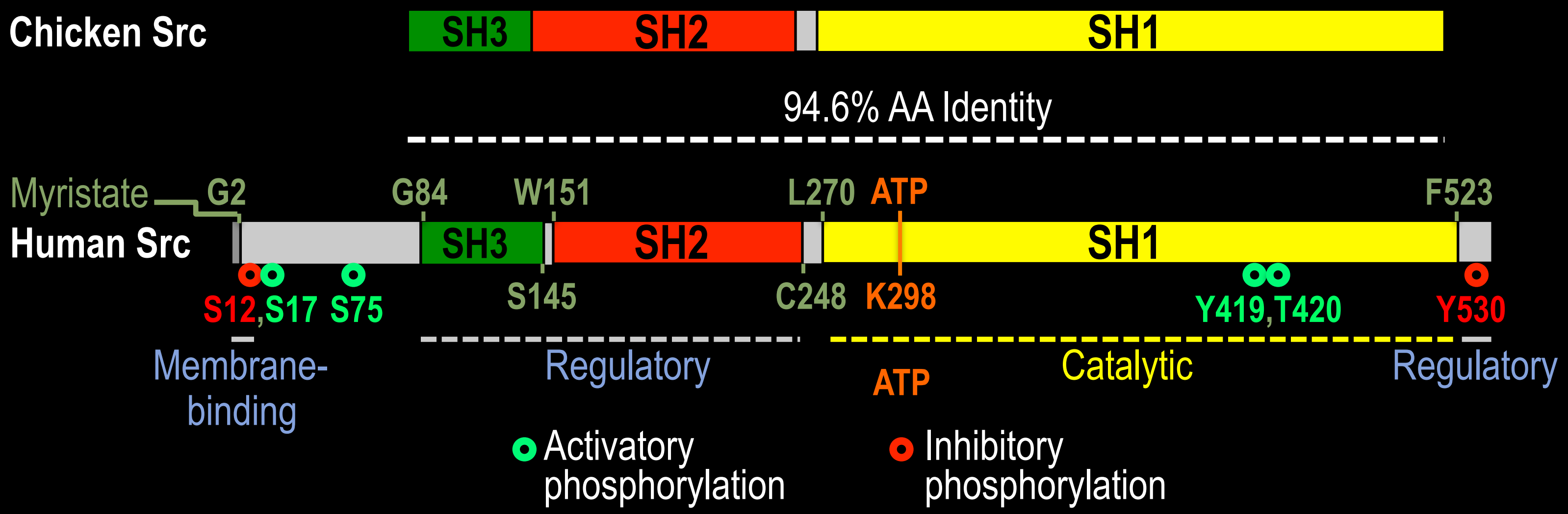
Protein Kinases and Oncology

- Importance of protein kinases in cancer was identified early in oncology research
- First oncogene indirectly discovered with the Rous Sarcoma Virus in 1910 by Francis Peyton Rous at Rockefeller University

Rous (1879-1970)

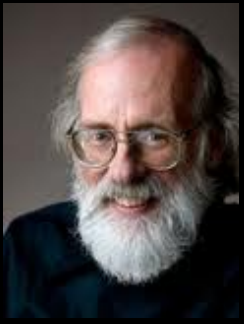
KINEXUS

Src Protein Structure



- Michael Bishop and Harold Varmus at the University of California, San Francisco determined the structure of the Src gene
- Tony Hunter at the Salk Institute discovered protein-tyrosine phosphorylation in Rous Sarcoma Virus infected cells

Bishop (1936-) Varmus (1939-)



Hunter (1943-)

Retrovirus Transduced Oncogenes

- Most oncogenes identified from retroviruses encode protein kinases and other associated cell signalling proteins

Name (including alias)	Protein Type	Retrovirus	Retrovirus Host
Abl1; Abl; JTK7; RP11-83J21.1-003	Kinase-Tyr	Abelson murine leukemia	Mouse
Akt1; PKB; RAC; hCG_96740	Kinase-Ser	AKT8	Mouse
BRCA1; RNF53	Transcr. factor	Mill House virus 2	Chicken
CBL; CBL2; RNF55	Ubiquitination factor	Cas NS-1 virus	Mouse
Crk	Adapter	Avian sarcoma virus CT10	Chicken
EGFR; ERBB1	Kinase-Tyr receptor	Avian erythroblastosis ES4	Chicken
Fes; Fps	Kinase-Tyr	Fujinami sarcoma; Snyder-Theilen feline sarcoma	Chicken; cat
Fgr; Src2	Kinase-Tyr	Gardner-Rasheed feline sarcoma	Cat
Fms	Kinase-Tyr receptor	McDonough feline sarcoma	Cat
Fos; G0S7	Transcr. factor	Mouse	Mouse
HRas; HRas1	G protein	Harvey murine sarcoma	Mouse
Jun	Transcr. factor	Avian sarcoma virus 17	Chicken
Kit	Kinase-Tyr receptor	Hardy-Zuckerman feline sarcoma	Cat
KRas; K-Ras; KRas2; RASK2; PRa1	G protein	Kirsten murine sarcoma	Mouse
MAF	Transcr. factor	AS42 sarcoma virus	Chicken
Mos	Kinase-Ser	Moloney murine sarcoma	Mouse
MPL; TPOR; RP1-92O14.1-002; hCG_23161	Cytokine receptor	Myeloproliferative leukemia	Mouse
Myb; c-myb, hCG_32380	Transcr. factor	Avian myeloblastosis E26	Chicken
Myc; BHLHE39	Transcr. factor	Myelocytomatosis 29	Chicken
PDGFB; PDGF2; Sis	Growth factor	Simian sarcoma	Whoolly monkey
Raf1; Raf	Kinase-Ser	3611 murine sarcoma	Mouse
Rel	Transcr. factor	Reticuloendotheliosis	Turkey
Ros1; RP1-179P9.1-003	Kinase-Tyr receptor	UR2	Chicken
Ski	Transcr. factor	SKV770	Chicken
Src; Src1	Kinase-Tyr	Rous sarcoma	Chicken
THRA; EAR7; ErbA1; NR1A1; THRA1; THRA2	Nuclear receptor	Avian erythroblastosis ES4	Chicken
Yes1; Yes	Kinase-Tyr	Y73/Esh sarcoma	Chicken

Viral Insertional Mutagenesis

- Viral integration into the genomes can also induce cancers through activation of host genes to over produce normal proteins

Oncogene Name (including alias)	Protein Type	Insertional Mutagenesis-Mutagen	Insertional Mutagenesis-Species
BMI1; PCGF4; RNF51; RP11-573G6.1-002	Transcr. factor	Moloney murine leukemia virus (MoMLV)	Mouse
CCND1; BCL1; PRAD1	Kinase activator	Friend murine leukemia virus (FMLV)	Mouse
CCND2	Kinase activator	Moloney murine leukemia virus (MoMLV)	Mouse
EGFR; ErbB1	Kinase-Tyr receptor	Avian leukosis virus	Chicken
FGF3; Int2	Growth factor	Mouse mammary tumour virus	Mouse
GM-CSF	Cytokine	Intracisternal A particle (IAP)	Mouse
IL2	Cytokine	Gibbon ape leukemia virus (GaLV)	Gibbon ape
IL3	Cytokine	Intracisternal A particle (IAP)	Mouse
Int-H; Int-5	Enzyme - Estrogen synth.	Mouse mammary tumour virus (MMTV)	Mouse
KRas; K-Ras; KRas2; RasK2; PRa1	G protein	Friend murine leukemia virus (FMLV)	Mouse
Lck	Kinase-Tyr receptor	Moloney murine leukemia virus (MoMLV)	Mouse
MAP3K8; Cot; ESTF; Tpl2	Kinase-Ser	Moloney murine leukemia virus (MoMLV)	Mouse
Mos	Kinase-Ser	Intracisternal A particle (IAP)	Mouse
Myc; BHLHE39	Transcr. factor	Avian leukosis virus; Feline leukemia virus	Chicken; Cat
NOTCH4; Int3	Receptor	Mouse mammary tumour virus	Mouse
NOV; CCN3; IGFBP9; NOVH	Growth factor	Avian leukosis virus	Chicken
Pim1	Kinase-Ser	Moloney murine leukemia virus (MoMLV)	Mouse
Pim2	Kinase-Ser	Moloney murine leukemia virus (MoMLV)	Mouse
TP53; p53	Transcr. factor	Moloney murine leukemia virus (MoMLV)	Mouse
Wnt1; Int1	Growth factor	Mouse mammary tumour virus	Mouse

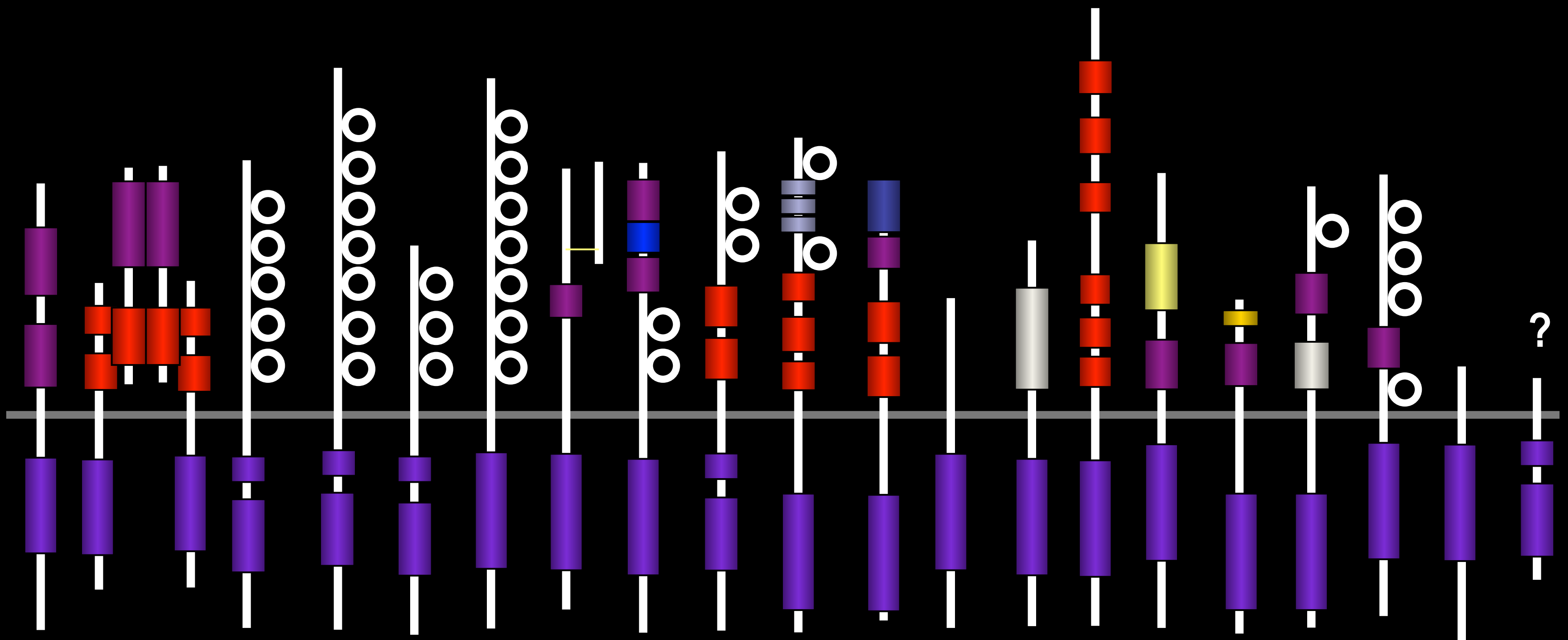
KiNEXUS

Amplified Oncogenes in Human Cancers

- Many of the oncogenes identified in retroviral studies with birds and mammals are also found in association with human cancers

Oncogene Name	Protein Type	Frequently Amplified Chromosomal Regions
AIB1; BTAK; NCOA3	Receptor	Breast cancers (15%)
Akt1; PKBα; RAC; hCG_96740	Kinase-Ser	Gastric cancers (20%)
Akt2; PKBβ	Kinase-Ser	Pancretic, ovarian cancers (30%)
CCND1; BCL1; PRAD1	Kinase activator	Breast and squamous cell carcinomas (40-50%)
CCNE1; CCNE	Kinase activator	Sarcomas (40%)
CDK4	Kinase-Ser	Sarcomas (40%)
CDK6	Kinase-Ser	Gliomas (5%)
EGFR; ErbB1; HER1	Kinase-Tyr receptor	Glioblastomas (50%); squamous cell carcinomas (10-20%)
ErbB2; HER2; MLN19; Neu; NGL	Kinase-Tyr receptor	Gastric, ovarian, breast carcinomas (10-30%)
ErbB3; HER3; c-erbB-3	Kinase-Tyr receptor	Oral squamous cell carcinomas
ErbB4; HER4	Kinase-Tyr receptor	Oral squamous cell carcinomas
FGFR1; FGFBR; FLG; FLT2	Kinase-Tyr receptor	Breast carcinomas (10%)
FGFR2	Kinase-Tyr receptor	Breast carcinomas
GLI1; GLI	Transcr. factor	Glioblastomas
KRas; K-Ras; KRAS2; RASK2; PRa1	G protein	Lung, ovarian, bladder carcinomas (5-10%)
K-SAM	Kinase-Tyr receptor	Gastric, breast carcinomas (10-25%)
Met	Kinase-Tyr receptor	Gastric carcinomas (20%)
Myb, c-Myb; hCG_32380	Transcr. factor	Colon carcinomas, leukemias
Myc; BHLHE39	Transcr. factor	Various leukemias and carcinomas (10-50%)
MycL1; hCG_15870	Transcr. factor	Lung carcinomas (10%)
MycN; BHLHE37; NMyc; hCG_1783900	Transcr. factor	Neuroblastomas, lung carcinomas (30%)
NRas; HRas1	G protein	Head and neck cancers (30%)

Structures of Receptor-tyrosine Kinases



EGFR
ErbB2
ErbB3
ErbB4

InsR
IGFR1
IRR

PDGFRα
PDGFRβ
CSF1R
Kit
Flk2

Flt1
Flk1
Flt4

FGFR1
FGFR2
FGFR3
FGFR4

CCK4

Met
Ron

TrkA
TrkB
TrkC

Axl
Ark
Eyk
Rse
Mer

Tie
Tek

EphA1
EphA2
EphA3
EphA4
EphA5
EphA6
EphA7
EphA8
EphA9
EphB1
EphB2
EphB3
EphB4
EphB5
EphB6

Ryk

DDR1
DDR2

Ros

Ret

Ltk
Alk

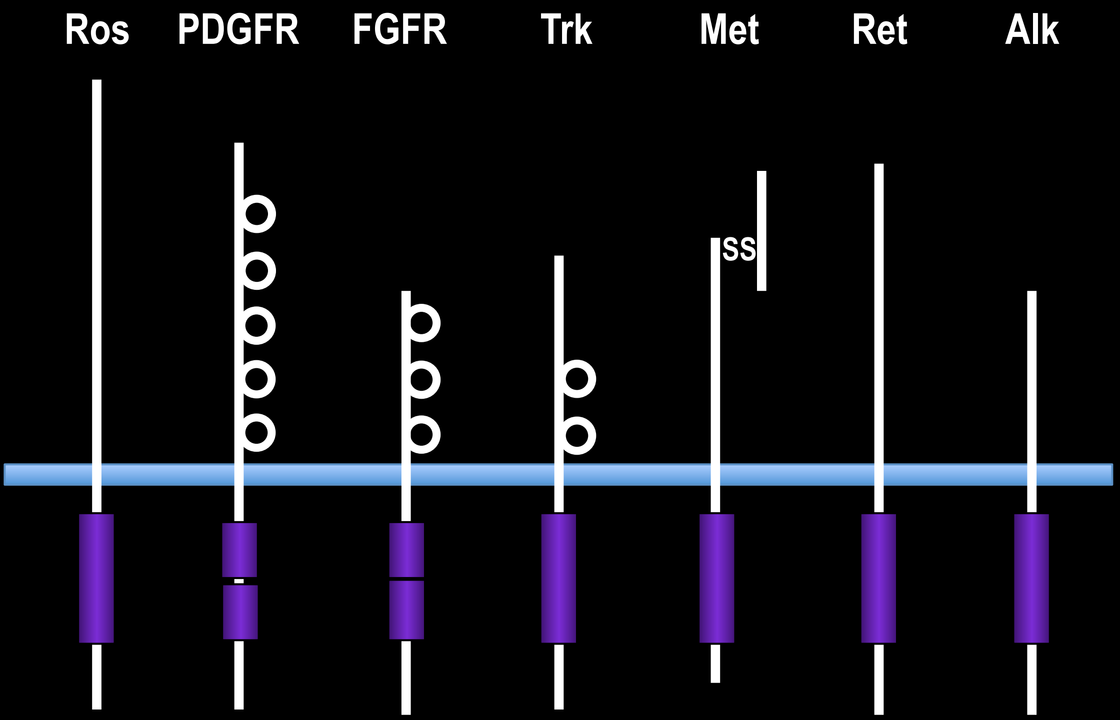
Ror1
Ror2

Musk

AATYK
AATYK2
AATYK3

RTK106

Fusions with Receptor-tyrosine Kinases



Ros
FIG

PDGFRα
BCR
PDGFRβ
Tel
H4
Hip1
Cev4
Rab5

FGFR1
FIM
FOP
CEP110
BCR
FGFR3
IgH
Tel

TrkA
Tpm3
Tpr
Tfg
TrkC
Tel

Met
Tpr

Ret
H4
R1a
ELE1
Htif1
Htifg
KTN1
RFG9
ELKS
PCM1

Alk
NPM

Receptor-Tyr kinases

Fusion partners

Gene Translocations and Cancer

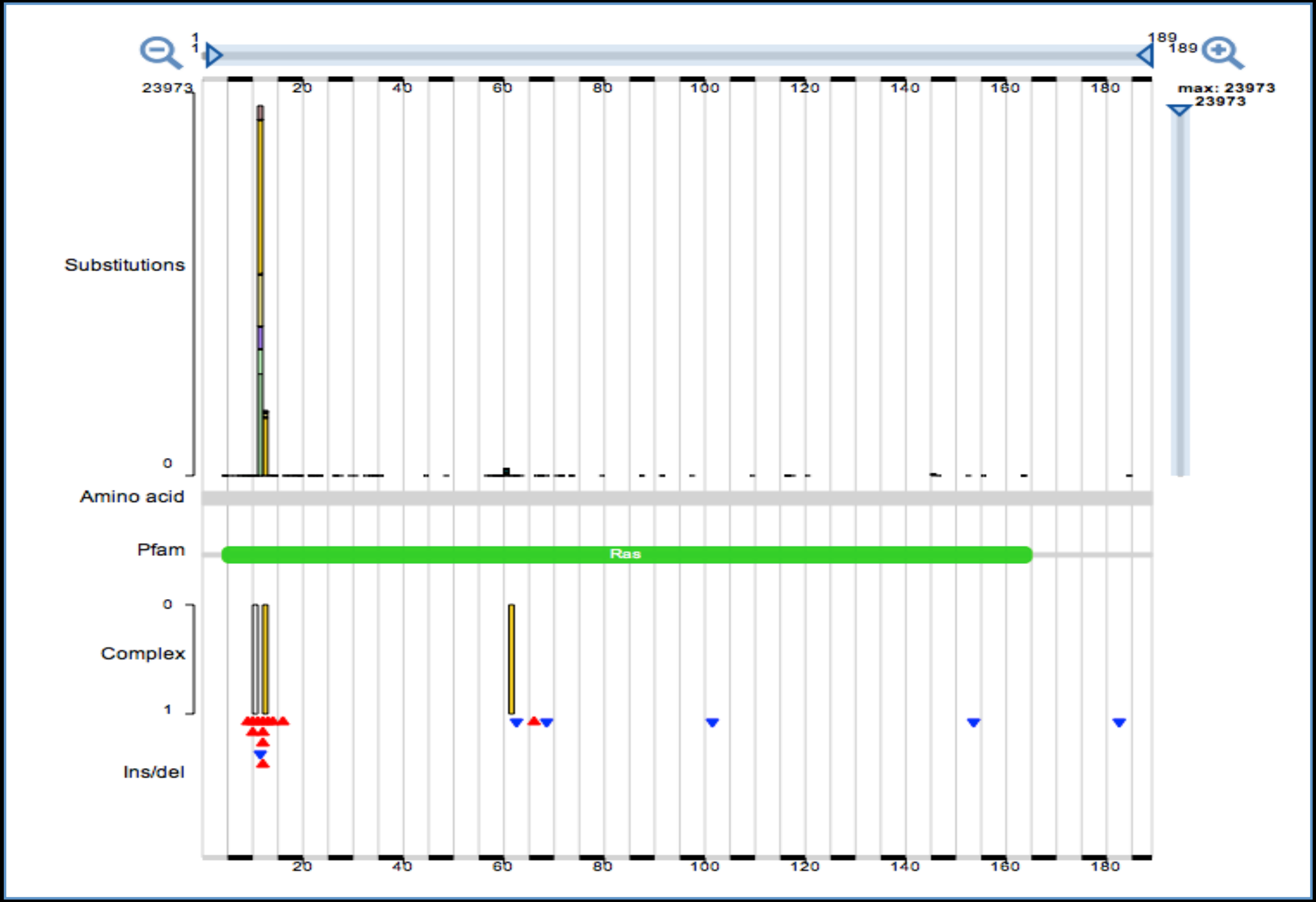
Oncogene Name (including alias)	Function/Activation	Translocations that deregulate and create oncogenes
ACTB	Cytoskeletal protein	Dermatofibrosarcoma protuberans with fusion with GLI1 promoter
AFF1; AF4; FEL; MLLT2; PBM1	Transcr. factor	Acute leukemias from hrx/af4 fusion
ALK	Kinase-Tyr receptor	Large-cell lymphomas from NPM/ALK fusion
Bcl2	Bcl protein	Follicular B-cell lymphomas
Bcl3; Bcl4; D19S37	Bcl protein	Chronic B-cell lymphomas
Bcl6; Bcl5; LAZ3; ZBTB27; ZNF51	Bcl protein	Diffuse B-cell lymphomas
Bcr-Abl	Kinase-Ser/Kinase-Tyr	Chronic myelogenous leukemias; acute lymphocytic leukemias with bcr/ab
DEK	Transcr. factor	Acute myeloid leukemia with dek/can fusion
ERG, p55	Transcr. factor	Myeloid leukemias with ?/erg fusion
EWS1/WT1	Transcr. factor	Ewing sarcomas with ews/fli fusion
FGFR1; FGFBR; Flg; FLT2	Kinase-Tyr receptor	Fusions with FIM, FOP, CEP110, and BCR
FGFR3; JTK4	Kinase-Tyr receptor	Multiple myelomas with fusions with IgH and Tel
HOX1	Transcr. factor	Acute T-cell leukemias
HOXA9; HOX1G	Transcr. factor	Acute myeloid leukemia from translocation t(7;11)(p15;p15) giving NUP98 fusion. Chronic myeloid leukemia from translocation t(7;17)(p15;q23) giving MSI2 fusion.
LMO1; RBTN1; RHOM1; TTG1	Transcr. factor	Acute T-cell leukemias
LYL1; BHLHA18	Transcr. factor	Acute T-cell leukemias
Met	Kinase-Tyr receptor	Fusions with Tpr
MLL; ALL1; CXXC7; HRX; HTRX; KMT2A; MLL1; TRX1	Transcr. factor	Acute leukemias from hrx/enl fusion; also at least 65 other partners
Myc; c-Myc; BHLHE39	Transcr. factor	Burkitt's lymphoma; other B- and T-cell malignancies
MYH11; KIAA0866	Transcr. factor/ Cytoskeletal protein	Acute myeloid leukemias with CBFbeta/MYH11 fusion
NFKB2; LYT10	Transcr. factor	B-cell lymphoma from lyt-10/Calpha1fusion
NPM1; NPM	Centrosome protein	Large-cell lymphomas from NPM/ALK fusion
NTRK1; TRK; TrkA	Kinase-Tyr receptor	Fusions with Tpm3, Tpr, and Tfg
NTRK3; TrkC	Kinase-Tyr receptor	Fusions with Tel
NUP214; CAIN; CAN; KIAA0023	Nuclear pore protein	Translocation t(6;9)(p23;q34) with DEK-CAN fusion. Translocation t(6;9)(q21;q34.1) with SET.
PBX1; PRL	Transcr. factor	Acute pre-B-cell leukemias with E2A/pbx1 fusion
PDGFRA	Kinase-Tyr receptor	Fusions with BCR
PDGFRB	Kinase-Tyr receptor	Fusions with Tel, H4, Hip1, Cev4, and Rab5
PML; MYL; RNF71; TRIM19	Transcr. factor	Acute promyelocytic leukemias from PML/RAR fusion
Ret; CDHF12	Kinase-Tyr receptor	Papillary thyroid carcinomas with fusions with H4, R1a, ELE1, Htif1, Htifg, KTN1, Sarcoma from FIG fusion
Ros1; RP1-179P9.1-003	Kinase-Tyr receptor	RFG9, ELKS, and PCM1
RUNX1; AML1; CBFA2	Transcr. factor	Acute myeloid leukemias with aml1/mtg8 fusion
RUNX1T1; AML1T1; CBFA2T1; CDR; ETO; MTG8; ZMYND2	Transcr. factor	Acute myeloid leukemia (AML-M2)from translocation t(8;21)(q22;q22) with RUNX1/AML1 fusion
TAL1; BHLHA17; SCL; TCL5	Transcr. factor	Acute T-cell leukemias
TAN1; NOTCH1	Transcr. factor	Acute T-cell leukemias

Definition of Oncoprotein

- Oncoproteins (OPs) are mutated in a highly focused manner such that the resultant protein has a gain of function
- Such mutations can:
  - Mimic an activating phosphorylation (e.g. B-Raf)
  - Inactivate an inhibitory site or region (e.g. K-Ras)
  - Produce a new product (e.g. IDH1)
  - Inhibit the degradation of the protein
- OPs are mutated narrowly across their genes and feature more point mutations and fewer complex mutations, insertions and deletions

OP Example - Sites of K-Ras Mutation

◆ G12 position is critical for binding and hydrolysis of GTP in Ras-related proteins; mutation leads to constitutive activation



From: http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=KRAS#histo

Mutations Locations in Protein Kinases

- Many of the stimulatory mutations in protein kinases occur with the “Kinase Activation” loop between kinase subdomains VII and VIII
- Phosphorylation of amino acid residues within the “Kinase Activation” loop typically activates most protein kinases and some substitution mutations mimic phosphorylation (e.g. B-Raf V600E)

Protein Name	Most Common Mutation Site	% Freq. Mutated in Tumours		VII		VII	VII	VII		VII								Gap 8												VIII	VIII	VIII
BRaf	V600E	18.8	K	I	G	D	F	G	L	A	T	V	K	S	R	W	S	4	F	E	Q	L	S	G	S	I	L	W	M	A	P	E
Ret	M918T	7.1	K	I	S	D	F	G	L	S	R	D	V	Y	E	E	D	4	R	S	Q	G	R	I	P	V	K	W	M	A	I	E
PDGFRA	D842V	3.2	K	I	C	D	F	G	L	A	R	D	I	M	H	D	S	4	K	G	S	T	F	L	P	V	K	W	M	A	P	E
EGFR	L858R	7.6	K	I	T	D	F	G	L	A	K	L	L	G	A	E	E	4	A	E	G	G	K	V	P	I	K	W	M	A	L	E
Met	Y1253D	0.8	K	V	A	D	F	G	L	A	R	D	M	Y	D	K	E	6	K	T	G	A	K	L	P	V	K	W	M	A	L	E
ALK	R1275Q	3.3	K	I	G	D	F	G	M	A	R	D	I	Y	R	A	S	4	G	G	C	A	M	L	P	V	K	W	M	P	P	E
Kit	D816V	4.9	K	I	C	D	F	G	L	A	R	D	I	K	N	D	S	4	K	G	N	A	R	L	P	V	K	W	M	A	P	E
Ft3	D835Y	0.5	K	I	C	D	F	G	L	A	R	D	I	M	S	D	S	4	R	G	N	A	R	L	P	V	K	W	M	A	P	E

Identification of Gene Mutations

- Over 60 million signal nucleotide polymorphisms appear to exist in human genomes
- Sequencing of genomes from tumours and comparison with data from normal tissues from the same patients is uncovering shared mutations in genes that encode proteins and microRNA
- Over 2500 genes appear to be abnormally mutated in diverse cancers
- The Sanger Institutes COSMIC Database tracks many of these known mutations: http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=bygene
- SNP chips permit rapid identification of SNPs and known mutations
- Cost of genome-wide sequencing (GWS) is approached $1000 per genome – more cancer-associated genes will be identified by GWS

Most Common Mutated Oncoproteins

- Most common oncoproteins are protein kinases and G proteins in human cancers

Gene Name	Uniprot ID	Protein Type	Viral Transduced	Gene Trans-location	# Mutations	# Specimens	% Mutated	% Mutation Per 100 AA	Most Common Mutations (number)
KRAS	P01116	G protein - Monomeric	YES	YES	26621	124694	21.4	11.3	G12D (9250); G12V (6190); G13D (3478); G12C (3173)
NPM1	P06748	RNA-binding		YES	4692	26112	18.0	6.1	W288fs* (Insertion - Frameshift)
JAK2	O60674	Kinase - Non-receptor-tyrosine kinase		YES	35015	100757	34.8	3.1	V617F (30,461)
NRAS	P01111	G protein - Monomeric			3167	55,047	5.8	3.0	Q61R (870); Q61K (663); G12D (404); Q61L (196)
IDH1	O75874	Metabolic enzyme - Isocitrate dehydrogenase			5507	45569	12.1	2.9	R132H (4164); R132C (620); R132S (116)
B-RAF	P15056	Kinase - Protein-serine kinase		YES	30889	153189	20.2	2.6	V600E (17,632); V600K (336)
MYD88	Q99836	Adapter/Scaffold			710	10389	6.8	2.2	L265P (659); S219C (13)
FLT3	P36888	Kinase - Receptor-tyrosine kinase			15006	81348	18.5	1.9	D835Y (164); D835? (695)
KIT	P10721	Kinase - Receptor-tyrosine kinase	YES		6384	35395	18.0	1.8	D816V (1221); V559D (197); D816? (188); V560D (160)
FGFR3	P22607	Kinase - Receptor-tyrosine kinase		YES	3038	20459	14.9	1.8	S249C (1637); Y373C (395); R248C (363); G370C (108)
EGFR	P00533	Kinase - Receptor-tyrosine kinase	YES		15952	81624	19.5	1.6	L858R (5,534); T790M (650)
HRAS	P01112	G protein - Monomeric	YES		930	32527	2.9	1.5	G12V (253); Q61R (134); Q61L (125); G13R (70)
CTNNB1	P35222	DNA binding - Transcription factor		YES	4092	36862	11.1	1.4	T41A (768); S45F (500); S37F (177); S33C (170)
PIK3CA	P42336	Lipid kinase			5823	50946	11.4	1.1	H1047E (2003); E545K (1146); E542K (721)
GNAS	P63092	G protein -Trimeric			476	13020	3.7	0.9	R201C (277); R201H (102); Q227L (20); Q227R (12)
FOXL2	P58012	DNA binding - Transcription factor			355	10843	3.3	0.9	C134W (338)
DNMT3A	Q9Y6K1	DNA methyltransferase			1030	13403	7.7	0.8	R882H (272), R282C (174)
GATA1	P15976	DNA binding - Transcription factor			296	8664	3.4	0.8	Most mutations in first 78 AA
IDH2	P48735	Metabolic enzyme - Isocitrate dehydrogenase			1168	33186	3.5	0.8	R140Q (738); R172K (269)
ABL1	P00519	Kinase - Non-receptor-tyrosine kinase	YES	YES	1185	13977	8.5	0.8	T315I (192); G250E (105); M244V (77); E255K (76)
GNAQ	P50148	G protein -Trimeric			327	12337	2.7	0.7	Q209P (143); Q209L (128)
PTPN11	Q06124	Phosphatase - Non-receptor-tyrosine phosphatase			534	15480	3.5	0.6	E76K (75); D61V (56); A72V (38); A72T (36)
TSHR	P16473	Receptor - G protein-coupled			390	8807	4.4	0.6	T632I (44); M453T (34); A623V (20); D619G (19)
GNA11	P29992	G protein -Trimeric			193	9743	2.0	0.6	Q209L (150); R183C (7)
PDGFRA	P16234	Kinase - Receptor-tyrosine kinase		YES	1442	24559	5.9	0.5	D842V (450); V561D (51); N659K (13)
RET	P07949	Kinase - Receptor-tyrosine kinase		YES	692	14282	4.9	0.4	M918T (379); C634R (20); A883F (10)
EZH2	Q15910	Histone-lysine N-methyltransferase			412	13102	3.1	0.4	Y646F (60); Y646N (51); Y646S (22); Y646H (19)
MPL	P40238	Receptor (for Thrombopoietin)	YES		614	23326	2.6	0.4	W515L (333); W515K (103)
NOTCH1	P46531	Receptor (for Jagged ligands)		YES	1250	15919	7.9	0.3	Most mutations between AA 1555-1784 and AA 2272-2529
PPP2R1A	P30153	Protein phosphatase regulator			143	9133	1.6	0.3	R183W (23); P179R (22); R183Q (10); S256F (12)

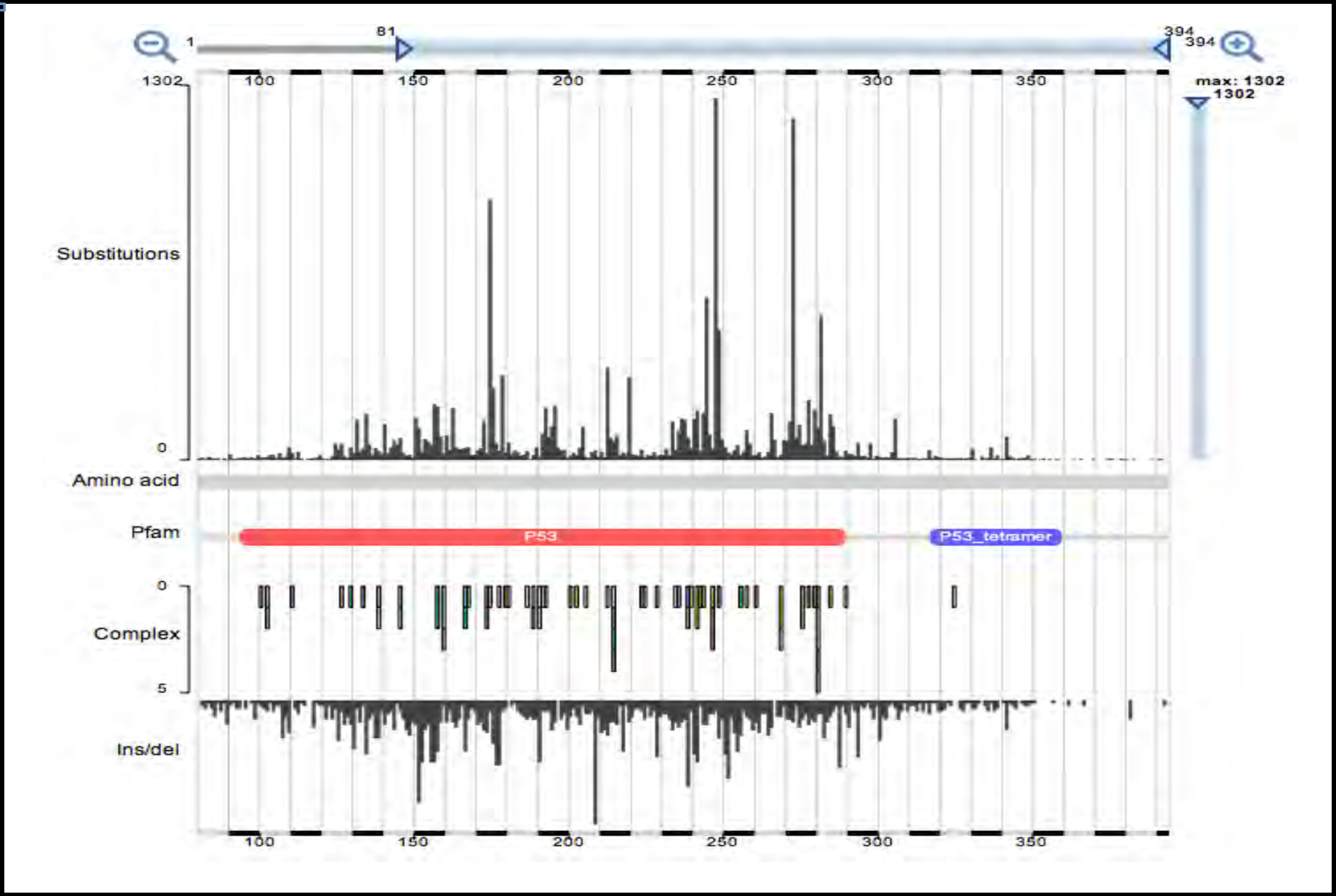
KINEXUS

Definition of a Tumour Suppressor Protein

- Tumour suppressor proteins (TSPs) are typically mutated broadly and feature many types of mutations as a general loss of function for these proteins drives cancer development
- Such mutations can:
  - Produce disruptions or deletions of functional regions:
    - Inactivate a catalytic region
    - Prevent binding to activators and downstream targets
  - Stimulate the degradation of the protein

TSP Example - Sites of p53 Mutation

- Mutational hot spots in p53 are commonly involved in DNA binding, which becomes disrupted



From: http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=TP53#histo

Most Common Tumour Suppressor Proteins

- Most common mutated TSPs in human cancers are transcription factors, protein phosphatases and adhesion proteins

Gene Name	Uniprot ID	Protein Type	Inherited	# Mutations	# Specimens	% Mutated	% Mutation Per 100 AA	Most Common Mutations (number)
CDKN2A	Q8N726	Protein kinase inhibitor	YES	4043	32333	12.5	8.0	R80* (88); R58* (72); W110* (30)
TP53	P04637	DNA binding - Transcription factor	YES	23881	81164	29.4	7.5	R175H (857); R248Q (614); R273H (560); R248W (547)
VHL	P40337	Protein synthesis	YES	1849	15168	12.2	5.7	S68* (17); L158V (13); L169P (12); L89H (12); S65L (11)
PTEN	P60484	Lipid phosphatase	YES	2726	28995	9.4	2.3	R130G (102); R130Q (66)
SMARCB1	Q12824	Regulator of chromatin remodelling	YES	471	9977	4.7	1.2	R201* (21)
NF2	Q16230	Regulator of Hippo/SWH signalling	YES	812	12134	6.7	1.1	R57* (28); R341* (21); R262* (12); R196* (12)
CEBPA	P49715	DNA binding - Transcription factor	YES	635	16184	3.9	1.1	Diverse and broad
SMAD4	Q13485	DNA binding - Transcription factor	YES	692	12966	5.3	1.0	R361H (20); R361C (16); D351H (10)
IKZF1	Q13422	DNA binding - Transcription factor		468	9189	5.1	1.0	Diverse and broad
RUNX1	Q01196	DNA binding - Transcription factor	YES	538	13001	4.1	0.9	Diverse and broad
WT1	P19544	DNA binding - Transcription factor		560	14957	3.7	0.8	S313* (12); R390* (9) ;D396N (9); R394W (7); R394Q (6)
LKB1	Q15831	Kinase - Protein-serine kinase	YES	381	13032	2.9	0.7	F354L (26); D194Y (8)
MEN1	O00255	DNA binding - Transcription factor	YES	401	10544	3.8	0.6	Diverse and broad
FBXW7	Q969H0	Proteolysis - Ubiquitination		655	15486	4.2	0.6	R465C (67); R465H (59); R505C (57); R479Q (33)
RB1	P06400	DNA binding - Transcription factor	YES	633	11804	5.4	0.6	Diverse and broad
PAX5	Q02548	DNA binding - Transcription factor		208	9346	2.2	0.6	P80R (22)
APC	P25054	Regulator of Wnt signalling	YES	3189	21050	15.2	0.5	R1450* (160)
CHK2	O96017	Kinase - Protein-serine kinase	YES	232	8527	2.7	0.5	K373E (52); A392A (52); Y390C (37); S372S (36); P536L (32)
ASXL1	Q8IXJ9	Regulator of transcription		927	12658	7.3	0.5	E1102D (16)
TET2	Q6N021	Metabolic enzyme - Methylcytosine dioxygenase		1143	14043	8.1	0.4	Diverse and broad
GATA3	P23771	DNA binding - Transcription factor		149	8394	1.8	0.4	Diverse and broad
BAP1	Q92560	Deubiquitinating enzyme	YES	254	8851	2.9	0.4	Diverse and broad
TNFAIP3	P21580	Protease - Ubiquitination		279	9125	3.1	0.4	Diverse and broad
MAP2K4	P45985	Kinase - Protein-dual specificity kinase		163	10570	1.5	0.4	Diverse and broad
CDH1	P12830	Adhesion - Cadherin	YES	297	8766	3.4	0.4	Diverse and broad
DKK2	Q9UBU2	Regulator of Wnt signalling		73	7606	1.0	0.4	C127R (7)
CDH10	Q9Y6N8	Adhesion - Cadherin		197	7213	2.7	0.3	Diverse and broad
PTCH1	Q13635	Receptor (for hedgehog ligands)	YES	542	10858	5.0	0.3	Diverse and broad
PBRM1	Q86U86	Regulator of chromatin remodelling		414	7884	5.3	0.3	Diverse and broad
FAM123B	Q5JTC6	DNA binding - Transcription factor		259	9707	2.7	0.3	R358* (8); R353* (7)

Most Commonly Mutated Cancer Genes in Tissues - 1

					Biliary tract	Blood & lymphoid	Bone	Breast	CNS	Cervix	Endometrium	Esophagus	Kidney	Large intestine	Liver	Lung	Meninges	Ovary	Pancreas	Parathyroid	Prostate	Salivary gland	Skin	Soft tissue	Stomach	Thyroid	Urinary tract
Name	UniProt ID	Protein Type	Type	AA	% Mutated/100 AA																						
CDKN2A	Q8N726	Kinase inhibitor	TSP	157	16	8	8	1	10	5	2	9	3	1	6	7	5	4	12	0	1	6	14	8	3	6	10
TP53	P04637	Transcr. factor	TSP	394	10	3	4	6	5	1	6	11	2	11	8	9	1	11	9	0	4	5	7	4	8	3	8
KRAS	P01116	G protein	OP	189	14	3	0	1	0	4	8	1	0	18	2	9	0	7	30	0	3	2	1	2	3	1	2
PTEN	P60484	Lipid phosphatase	TSP	404	2	1	1	1	4	1	10	0	1	3	1	1	0	1	0	0	3	1	3	1	1	1	1
NRAS	P01111	G protein	OP	190	1	5	0	0	0	1	1	0	0	2	1	0	1	0	0	0	0	0	8	2	1	3	1
VHL	P40337	Protein synthesis	TSP	214	0	0	0	0	0	0	1	0	18	5	0	0	0	0	2	0	0	0	0	7	0	2	0
HRAS	P01112	G protein	OP	190	0	0	1	0	0	4	0	0	0	0	0	0	0	0	0	0	2	6	4	2	1	2	5
CTNNB1	P35222	Transcr. factor	OP	782	1	0	1	0	1	1	3	0	1	1	3	0	0	1	2	0	1	1	1	6	1	2	0
B-RAF	P15056	Kinase-Ser	OP	767	1	1	0	0	1	0	0	0	0	2	0	0	0	1	0	0	0	0	6	0	0	5	0
IDH1	O75874	Metabolic enzyme	OP	415	2	1	6	0	8	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	2	0
SMARCB1	Q12824	Chromatin	TSP	386	0	0	4	0	4	0	1	0	0	1	0	0	1	0	0	0	0	0	0	11	0	0	0
IFITM3	Q01628	Antiviral	TSP	133	0	0	0	0	1	11	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	1
NF2	P35240	GTPase	TSP	596	0	0	1	0	0	1	1	0	0	1	0	0	5	0	0	0	0	0	0	7	0	1	0
GNAS	P63092	GTP exchange	OP	395	6	0	0	0	0	0	1	0	0	1	0	0	0	1	2	2	0	0	0	2	0	1	0
SMAD4	Q13485	Transcr. factor	TSP	533	3	0	0	0	0	0	1	1	0	3	0	0	0	0	3	0	0	0	0	0	1	3	0
PIK3CA	P42336	Lipid kinase	OP	1069	1	0	0	2	0	1	2	1	0	1	0	0	0	1	0	0	0	0	1	0	1	0	2
AXIN1	O15169	Scaffold	TSP	863	3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	5	0
SEMA4D	Q92854	Receptor	TSP	123	0	0	0	0	0	0	3	1	0	3	0	1	0	0	0	0	1	0	0	0	0	0	1
IFITM1	P13164	Antiviral	TSP	126	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FBXW7	Q969H0	Ubiquitin conjug.	TSP	708	3	1	0	0	0	1	2	0	0	2	0	0	0	0	0	0	0	0	0	0	1	0	1
KIT	P10721	Kinase-Tyr	OP	977	0	2	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	1	5	0	0	0
RB1	P06400	Scaffold	TSP	929	0	0	2	0	1	0	1	0	0	2	0	1	0	0	0	0	0	1	0	0	0	0	2
FGFR3	P22607	Kinase-Tyr	OP	807	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	5
MEN1	O00255	Transcr. factor	TSP	611	0	0	0	0	0	0	1	0	0	1	0	0	0	0	2	4	0	0	0	1	0	0	0
NPM1	P06748	RNA-binding	OP	295	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LKB1	Q15831	Kinase-Ser	TSP	434	1	0	0	0	0	3	0	0	0	1	0	2	0	0	0	0	0	0	1	0	0	0	0
NFE2L2	Q16236	Transcr. factor	OP	606	0	0	0	0	0	2	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1
CDH10	Q9Y6N8	Adhesion	TSP	789	0	0	0	0	0	1	1	1	0	1	0	2	0	0	0	0	0	0	0	0	1	0	0
VEGFC	P49767	Cytokine	TSP	421	2	0	0	0	0	2	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
GATA1	P15976	Transcr. factor	OP	414	0	4	0	0	0	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

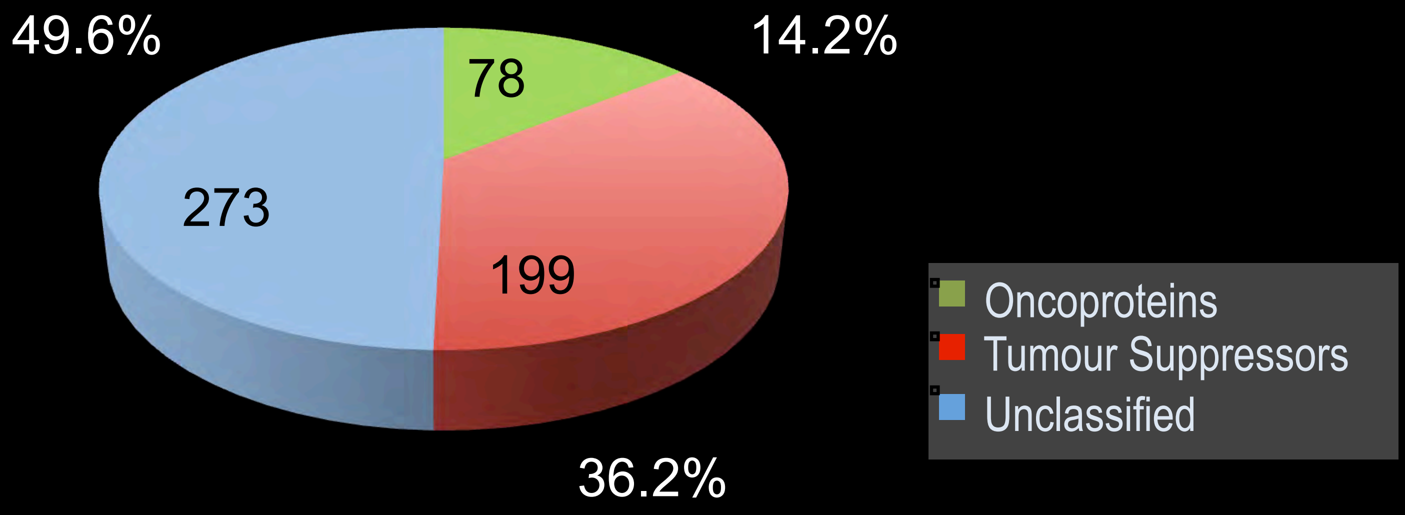
Most Commonly Mutated Cancer Genes in Tissues - 2

					Biliary tract	Blood & lymphoid	Bone	Breast	CNS	Cervix	Endometrium	Esophagus	Kidney	Large intestine	Liver	Lung	Meninges	Ovary	Pancreas	Parathyroid	Prostate	Salivary gland	Skin	Soft tissue	Stomach	Thyroid	Urinary tract
Name	UniProt ID	Protein Type	Type	AA	% Mutated/100 AA																						
CDH1	P12830	Adhesion	TSP	883	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	1
CHK2	O96017	Kinase-Ser	TSP	544	0	0	0	0	2	1	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1
DKK2	Q9UBU2	Wnt inhibitor	TSP	260	0	0	0	0	0	0	1	1	0	2	0	1	0	0	0	0	0	0	0	0	1	0	1
PIK3R1	P27986	Lipid kinase	OP	725	0	0	0	0	0	1	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
OPCML	Q14982	Adhesion	TSP	346	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	0	0	0	0	1	0	2
MKRN3	Q13064	Ubiquitin conjug.	TSP	508	0	0	0	0	0	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1
FAM113B	Q96HM7	Unknown	?	433	0	0	3	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
CDC73	Q6P1J9	Transcr. factor	TSP	532	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0
TSHR	P16473	Receptor-GPC	OP	765	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	4	0
MYD88	Q99836	Adapter	OP	310	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MAP2K4	P45985	Kinase-Ser	TSP	400	2	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
ACTR2A	P27037	Kinase-Ser	TSP	514	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	2	0	0
CDKN1A	P38936	Kinase inhibitor	TSP	165	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	5
CDC27	P30260	Regulatory	TSP	825	0	0	0	0	0	3	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
GNAQ	P50148	G protein	OP	360	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0
PPP2R1A	P30153	Phosphatase	OP	590	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FOXL2	P58012	Transcr. factor	OP	377	0	0	0	0	0	1	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
PEG3	Q9GZU2	Regulatory	TSP	1589	2	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
C15orf2	Q9NZP6	Regulatory	TSP	1157	0	0	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
CRIPAK	Q8N1N5	Kinase inhibitor	TSP	447	0	0	0	0	0	0	0	1	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0

- It is evident that different types of human cancers show preferences for different mutated proteins
- Some tissues show higher rates of mutations of proteins in cancer

Most Common Cancer-Associated Genes

Distribution of 550 Most Commonly Mutated Genes in Human Cancers



For Oncoproteins (OP)

Functional effect of mutation - Gain of function, promotes transformation = 71
Functional effect of mutation - Unclear effect on function, promotes transformation = 2
Functional effect of mutation - Unclear or mixed effect on function, may also be a TSP = 5

For Tumour Suppressor Proteins (TSP)

Functional effect of mutation - Loss of function, promotes transformation = 140
Functional effect of mutation - Unclear effect = 46
Functional effect of mutation - Unclear effect or mixed on function, and may also be an OP = 13

KiNEXUS

Distribution of OP and TSP by Protein Type

Distribution of 337 Most Commonly Mutated OP and TSP

Protein Type	Number	Percent
Adapter/scaffold	19	5.6
Adhesion	23	6.8
Cell cycle	5	1.5
Checkpoint	5	1.5
Chromatin	4	1.2
Cytoskeletal	12	3.6
DNA repair	3	0.9
Extracellular matrix	5	1.5
Extracellular mediator	1	0.3
G protein	6	1.8
GTPase activating protein	3	0.9
Ion channel	2	0.6
Metabolic enzyme	13	3.9
Proteolysis protein	10	3.0
Protein kinase-serine	19	5.6
Protein kinase-tyrosine	35	10.4
Protein phosphatase	9	2.7
Receptors	12	3.6
Regulatory	34	10.1
Transcription factor	47	13.9
Transport	8	2.4
Others	62	18.4

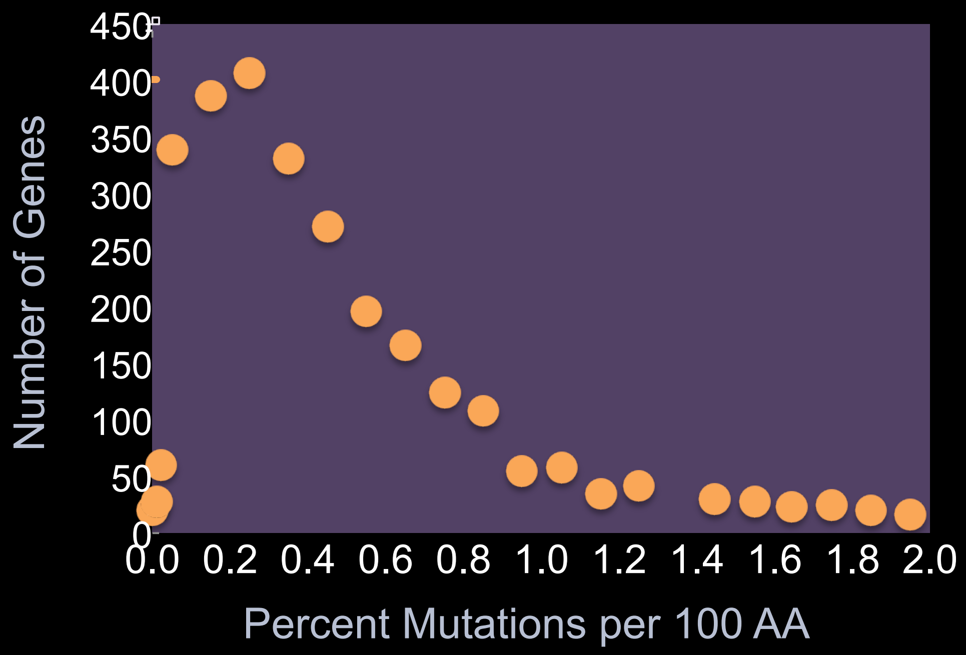
KINEXUS Human Cancer Protein KnowledgeBase

www.onconet.ca

- Open-access human cancer gene knowledgebase
- Beta-version currently available on-line
- Features mRNA expression and mutation information on 3000 proteins previously linked to cancer
- Classifies oncoproteins (OPs), tumour suppressor proteins (TSPs), and tumour-requiring proteins (TRPs)

Rate of Mutations in Human Cancers

Background Rates of Somatic Mutations



- Based on 2912 genes
- 176 entries including non-kinases had % Mutation per 100 amino acids (AA) rates less than 0.025
- 85 entries had % Mutation per 100 amino acids (AA) rates less than 0.01

	# Amino Acids	# Mutations	# Specimens	% Mutated	% Mutation Rate Per 100 AA
Average	773	117	7675	0.73	0.138
Median	403	29	7175	0.41	0.076

- Mutation is a random process, but natural selection in cancer favours the enrichment of gain- and loss-of-function mutations

KINEXUS

Definition of a Tumour Requiring Protein

- Tumour requiring proteins (TRPs) are “resistant” to mutations in cancer cells and may be required for the downstream actions of OP
- Rates of mutation in TRPs may be 10-fold lower than background mutation of proteins in general in cancer cells
- 72 proteins had only 0-2 mutations reported in over 7000 human tumours
- Many TRPs are protein kinases

KINEXUS

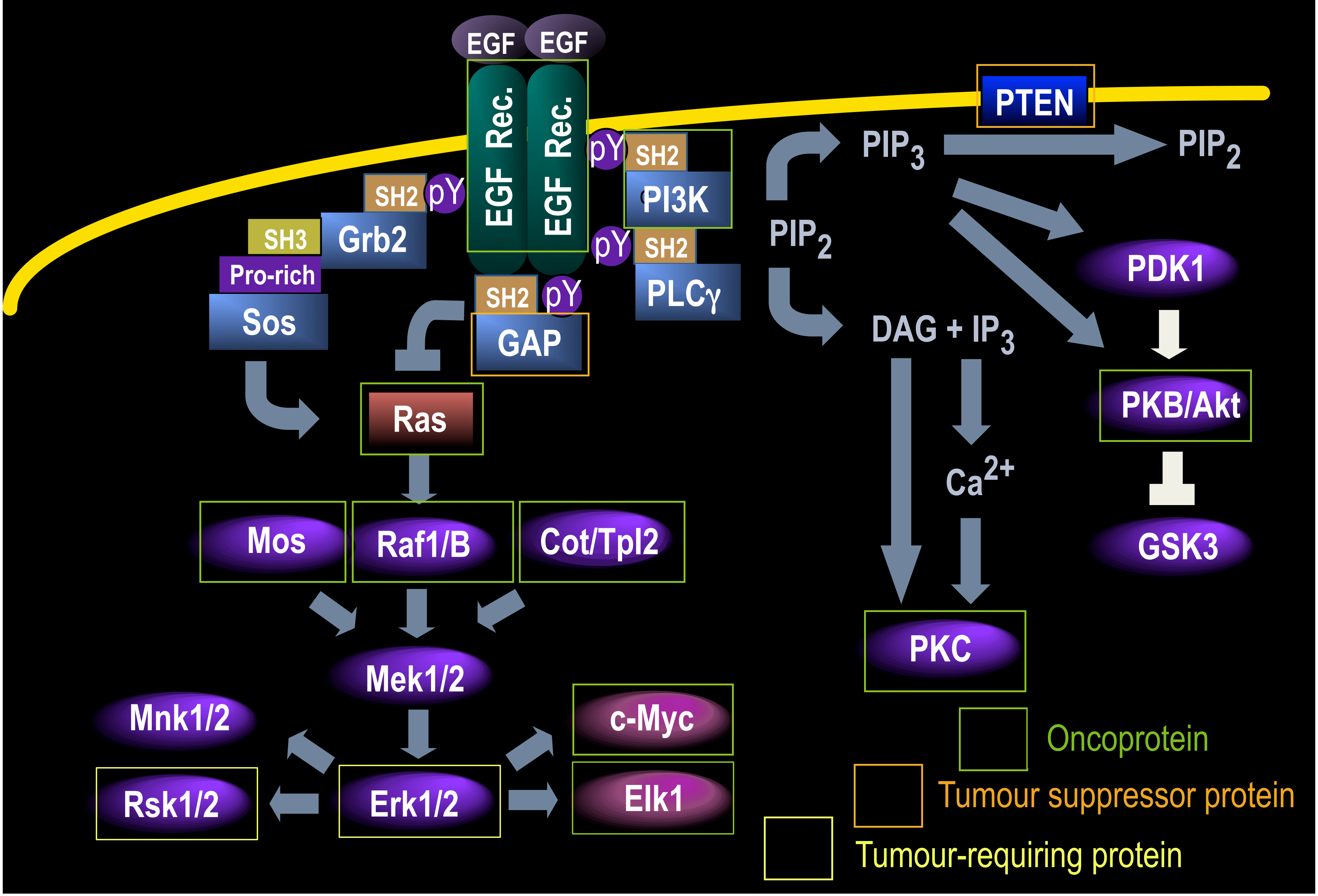
Tumour-Requiring Protein Kinases

- Many protein kinases are up and down regulated in human cancers, but appear to be “resistant” to mutation

Gene Name	Uniprot ID	Protein Type	Expression in Cancer	# Mutations	# Specimens	% Mutated	% Mutation Per 100 AA	Most Common Mutations (number)
PIM3	Q86V86	Kinase - Protein-serine kinase	Up	0	7473	0.00	0.000	No mutations reported
CDK20	A2A390	Kinase - Protein-serine kinase	Up	1	7498	0.01	0.003	Single point, missense mutation (A165V)
CDK10	Q15131	Kinase - Protein-serine kinase	Up	1	7459	0.01	0.004	Single point, missense mutation (N157S) in kinase catalytic domain
CDK1	P06493	Kinase - Protein-serine kinase	Up	1	7310	0.01	0.005	Single point, missense mutation (D73H)
MINK1	Q8N4C8	Kinase - Protein-serine kinase	Up	5	7894	0.06	0.005	4 point mutations and 1 deletion
MAPKAPK3	Q16644	Kinase - Protein-serine kinase	Up,Down	2	8562	0.02	0.006	2 point mutations (none were silent)
LATS1	O95835	Kinase - Protein-serine kinase	Down	6	7944	0.08	0.007	6 point mutations (none were silent)
RYK	P34925	Kinase - Receptor-tyrosine kinase	Up	4	8329	0.05	0.008	4 point mutations (none were silent)
ASK2	O95382	Kinase - Protein-serine kinase	Down	8	8473	0.09	0.008	7 point mutations (1 was silent) and 1 insertion
MAST2	Q6P0Q8	Kinase - Protein-serine kinase	Up	11	7529	0.15	0.008	11 point mutations
JNK3	P53779	Kinase - Protein-serine kinase	Up	3	8354	0.04	0.008	3 point mutations (1 was silent)
CSK	P41240	Kinase - Non-receptor-tyrosine kinase	Up	5	7462	0.07	0.015	5 point mutations (2 were silent)
PIM2	Q9P1W9	Kinase - Protein-serine kinase	Up	4	7930	0.05	0.015	2 point mutations, 1 deletion and 1 insertion
IKKg	Q9Y6K9	Kinase - Protein-serine kinase	Up	5	7584	0.07	0.016	3 point mutations (2 were silent), 1 insertion and 1 deletion
ERK4	P31152	Kinase - Protein-serine kinase	Up	8	8354	0.10	0.017	8 point mutations (1 was silent)
RSK1	Q15418	Kinase - Protein-serine kinase	Up	13	7462	0.17	0.024	13 point mutations (2 were silent)
MEKK2	Q9Y2U5	Kinase - Protein-serine kinase	Up	19	8353	0.23	0.028	17 point mutations, 1 deletion, and 1 insertion
FES	P07332	Kinase - Non-receptor-tyrosine kinase	Up	18	7498	0.24	0.029	18 point mutations
AMPK1	Q13131	Kinase - Protein-serine kinase	Up,Down	13	7931	0.16	0.029	13 point mutations (3 were silent)
ERK1	P27361	Kinase - Protein-serine kinase	Up	11	8354	0.13	0.035	11 point mutations (3 were silent)
AurB	Q96GD4	Kinase - Protein-serine kinase	Up	11	8439	0.13	0.038	11 point mutations (2 were silent)
CDK7	P50613	Kinase - Protein-serine kinase	Up	11	7916	0.14	0.040	11 point mutations
p38b	Q15759	Kinase - Protein-serine kinase	Down	14	8354	0.17	0.046	13 point mutations and 1 deletion
TP53RK	Q96S44	Kinase - Protein-serine kinase	Down	10	7285	0.14	0.054	8 point mutations (1 was silent), 1 insertion, and 1 deletion
ERK3	P53778	Kinase - Protein-serine kinase	Up,Down	17	8354	0.20	0.055	16 point mutations, and 1 deletion
MEK2	P36507	Kinase - Protein dual specificity kinase	Up	17	7660	0.22	0.055	16 point mutations, and 1 deletion
CDK6	Q00534	Kinase - Protein-serine kinase	Up,Down	17	8442	0.20	0.062	17 point mutations
ERBB3	P21860	Kinase - Receptor-tyrosine kinase	Up	71	8354	0.85	0.063	70 point mutations, and 1 deletion
CDK5	Q00535	Kinase - Protein-serine kinase	Up	17	7498	0.23	0.077	17 point mutations

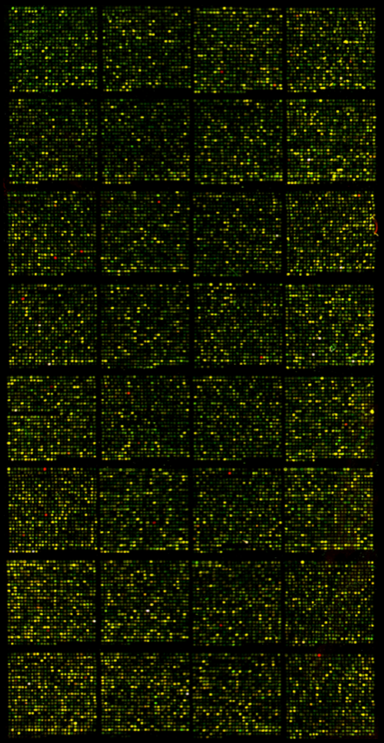
KINEXUS

Epidermal Growth Factor Signalling



TranscriptoNET KnowledgeBase

www.transcriptonet.ca



- Gene microarrays permit measurement of mRNA levels
- NCBI Gene Expression Omnibus (GEO) Database features gene chip data from >50,000 studies
- Kinexus developed TranscriptoNET to consolidate and normalize human gene expression data from 6000 normal and tumour cells and tissues to permit comparisons across studies

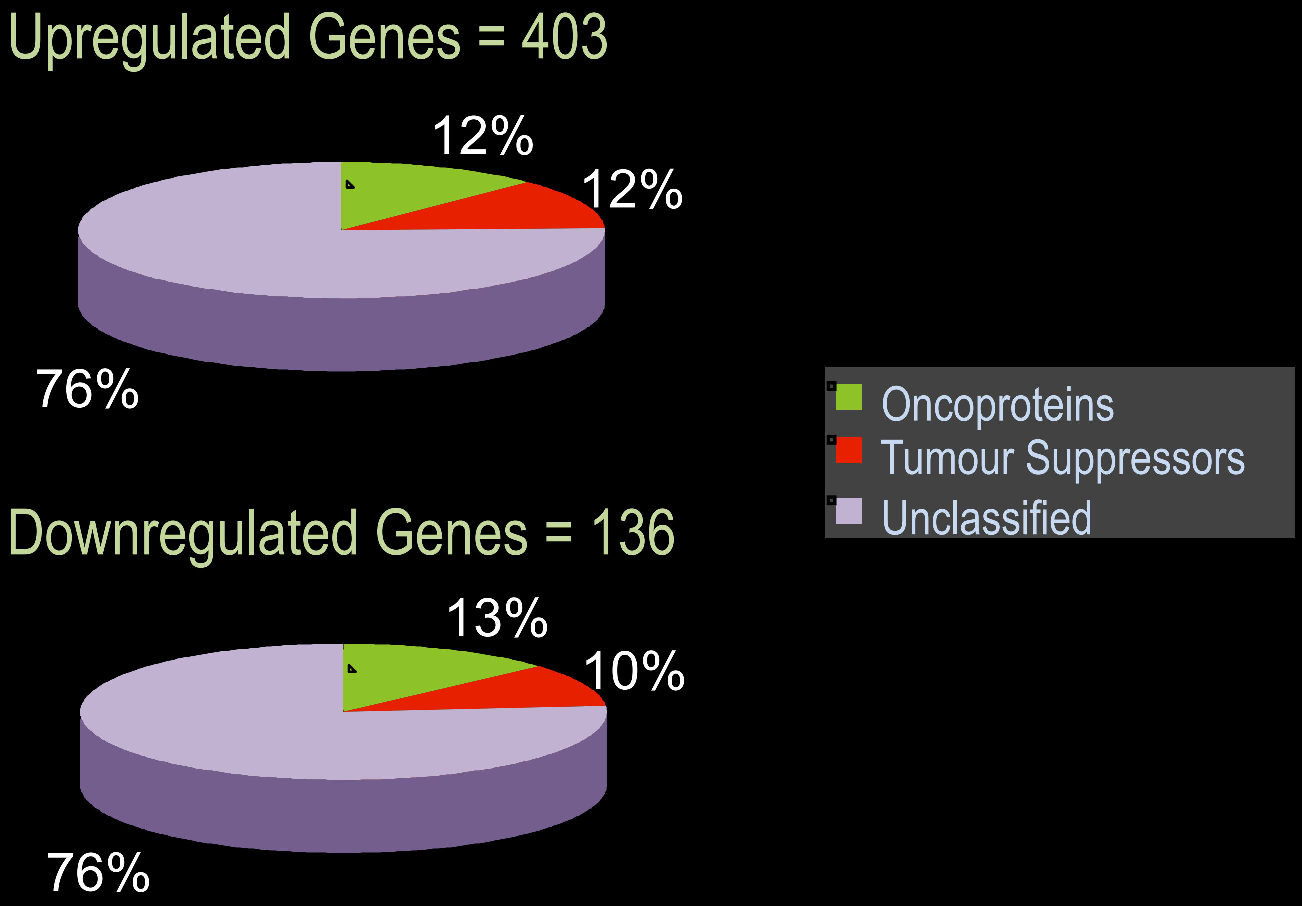
TranscriptoNET KnowledgeBase

www.transcriptonet.ca

- Open access human mRNA expression knowledgebase
- >600 human control and cancer tissues and cell lines
- ~21,000 human genes covered
- Contains data from over 125 million measurements
- Identifies highly, commonly and uniquely expressed genes in human tissues and cells
- Identifies differentially expressed genes in about 50 types of human cancers
- Identifies genes that are highly co-expressed
- Shows tissue and cell relatedness

Gene Expression in Human Cancers

- Many oncogenes are down-regulated and tumour suppressor genes up-regulated in human cancers


Upregulated Genes = 403

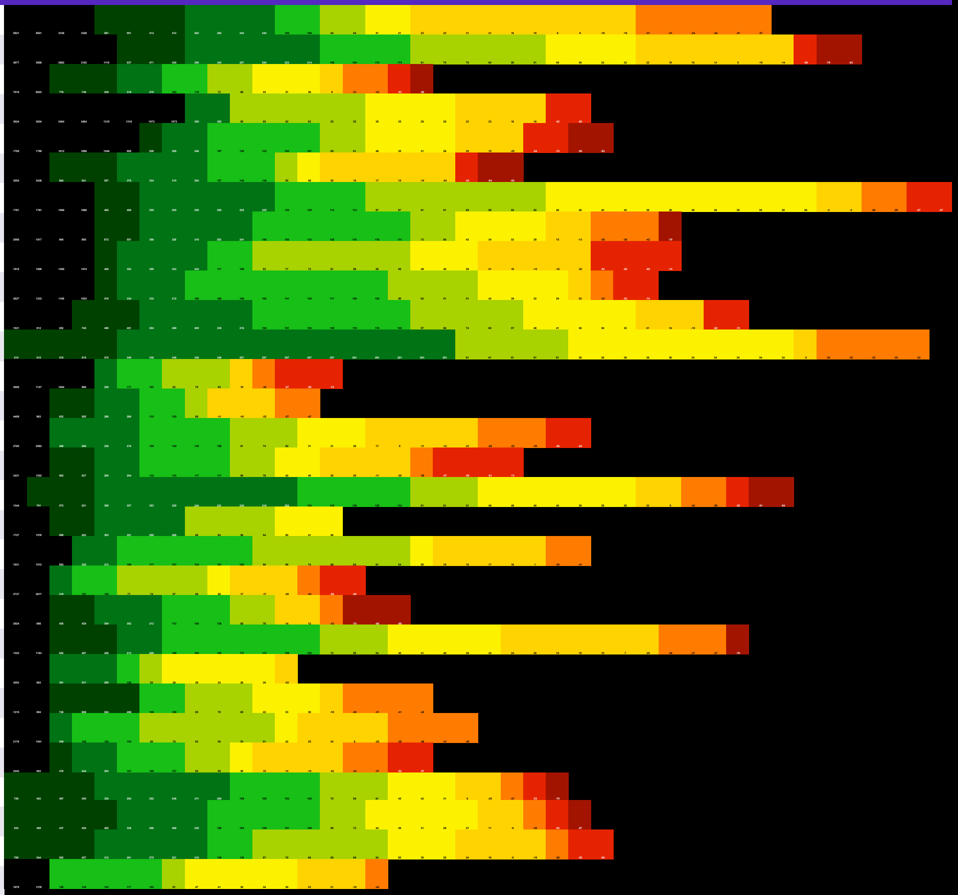
Downregulated Genes = 136

KINEXUS

Up-regulated Genes in Human Cancers

	Gene Name	Uniprot ID	Protein Type
Unlisted	SYTL2	Q6TV07	Regulator of G protein
OP	MMP9	P14780	Protease - matrix metalloproteinase
OP + TSP	ERBB3	P21860	Kinase - Receptor-tyrosine kinase
Unlisted	RAB13	P51153	G protein
TRP	CXCL10	P02778	Chemotactic factor
Unlisted	MMP12	P39900	Protease - matrix metalloproteinase
Unlisted	GAPDH	P04406	Metabolic enzyme - Dehydrogenase
Unlisted	RRM2	P31350	Metabolic enzyme - Reductase
Unlisted	C1QB	P02746	Complement
TSP?	ZWINT	O95229	Kinetochore
TRP	CKS2	P33552	Kinase regulatory subunit
Unlisted	EEF1D	P29692	Protein synthesis
TRP	S100A7	P31151	Calcium binding protein
Unlisted	PPA1	Q15181	Enzyme - Pyrophosphatase
TSP?	LY6E	Q16553	Receptor
Unlisted	TUBB2A	Q13885	Cytoskeletal
Unlisted	ISG15	P05161	Ubiquitin-like
TRP	TCEB2	Q15370	Protein synthesis
Unlisted	LAGE3	Q14657	
Unlisted	EIF2S3	P41091	Protein synthesis
OP	MMP1	P03956	Protease - Matrix metalloproteinase
Unlisted	MTHFD2	P13995	Metabolic enzyme - Dehydrogenase
Unlisted	OAZ1	P54368	Metabolic enzyme - Ornithine decarboxylase
Unlisted	UQCRQ	O14949	Cytochrome
Unlisted	YWHAZ	P63104	14-3-3.
TRP?	CORO1A	P31146	Cytoskeletal regulatory protein
Unlisted	MARCKSL1	P49006	Regulator
Unlisted	IFI6	P09912	Anti-apoptosis
TRP	PTTG1	O95997	Regulator of chromosome stability
Unlisted	MBOAT7	Q7L5N2	Metabolic enzyme - Acyltransferase

Expression in Cancer

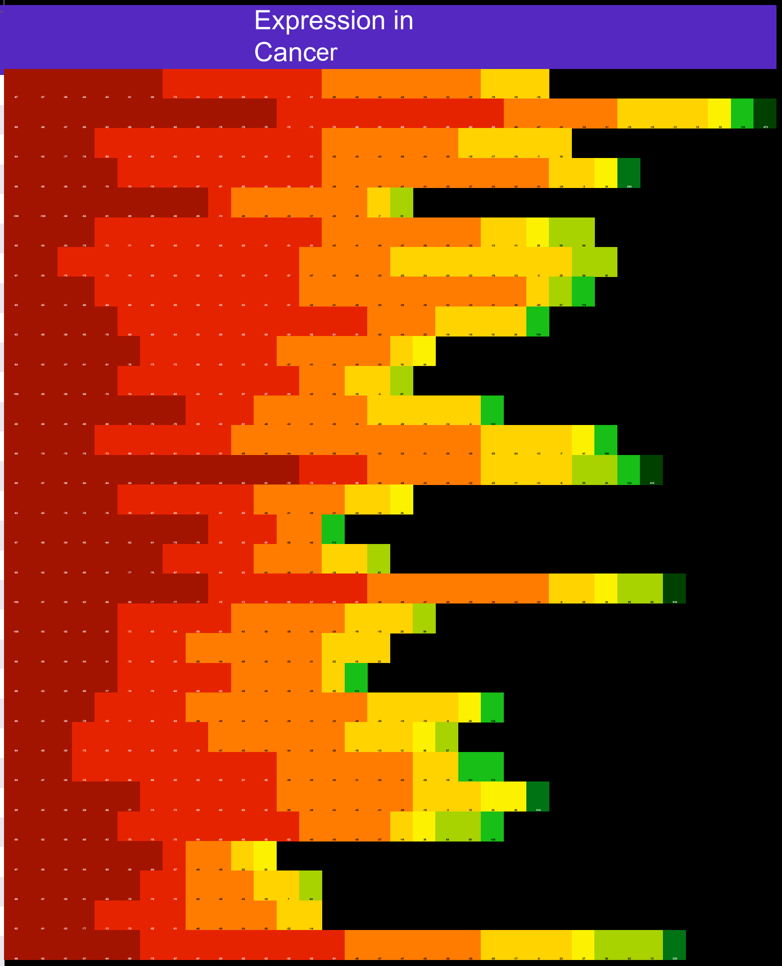


- ◆ Each tissue type per row is ordered with the highest increases in expression ordered from the left starting with black and descending

KINEXUS

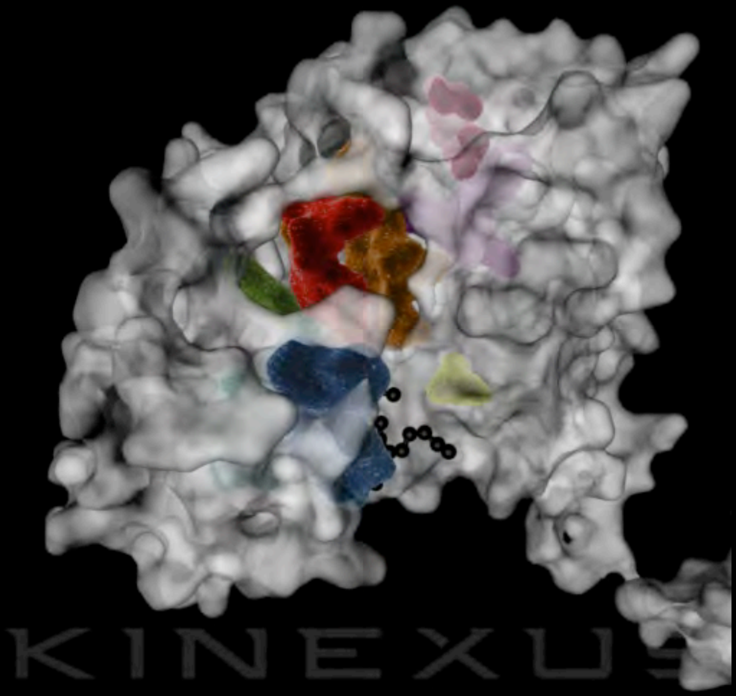
Down-regulated Genes in Human Cancers

	Gene Name	Uniprot ID	Protein Type
TRP	DPT	Q07507	Cell adhesion
Unlisted	ADH1B	P00325	Metabolic enzyme - Alcohol dehydrogenase
TRP	HLF	Q16534	DNA binding - Transcription factor
Unlisted	CLEC3B	P05452	Secretion
TSP	TIMP4	Q99727	Protease inhibitor
Unlisted	PDLIM3	Q53GG5	Cytoskeletal
Unlisted	C7	P10643	Immune
Unlisted	SYNM	O15061	Cytoskeletal
Unlisted	MAOB	P27338	Metabolic enzyme - Amine oxidase
Unlisted	ATP1A2	P50993	Transport - Sodium
Unlisted	LYVE1	Q9Y5Y7	Transport - Organelles
Unlisted	PCOLCE2	Q9UKZ9	Cell adhesion
TSP?	NR3C2	P08235	Receptor - Transcription factor
Unlisted	LTF	P02788	Transport - Iron
Unlisted	C14orf132	Q9NPU4	Uncharacterized
Unlisted	GABRP	O00591	Ion channel receptor
Unlisted	DES	P17661	Cytoskeletal
Unlisted	ABCA8	A1L3U3	Uncharacterized
Unlisted	EDN3	P14138	Vasoconstrictor peptide
Unlisted	PCK1	P35558	Metabolic enzyme - PEP carboxykinase
TRP	WISP2	O76076	Regulator of cell adhesion
Unclear	TXNIP	Q9H3M7	Transcription repressor
Unlisted	FCER1A	P12319	Immune
Unlisted	PALMD	Q9NP74	Regulator of cell shape
Unlisted	SFRP1	Q8N474	Ligand - for Wnt signalling
Unlisted	CLDN8	P56748	Tight junction - Claudin
Unlisted	PLIN1	O60240	Regulator for lipid metabolism
Unlisted	ADIPOQ	Q15848	Regulator for lipid metabolism
Unlisted	RERGL	Q9H628	GTPase
Unlisted	MFAP4	P55083	Cell adhesion



- Each tissue type per row is ordered with the highest decreases in expression ordered from the left starting with red and descending

Approved Kinase Drugs and Those in Development



EGF receptor 3D structure with kinase inhibitor drug Tarceva

- >19 drugs approved
- >150 in clinical trials
- >500 in pre-clinical trials

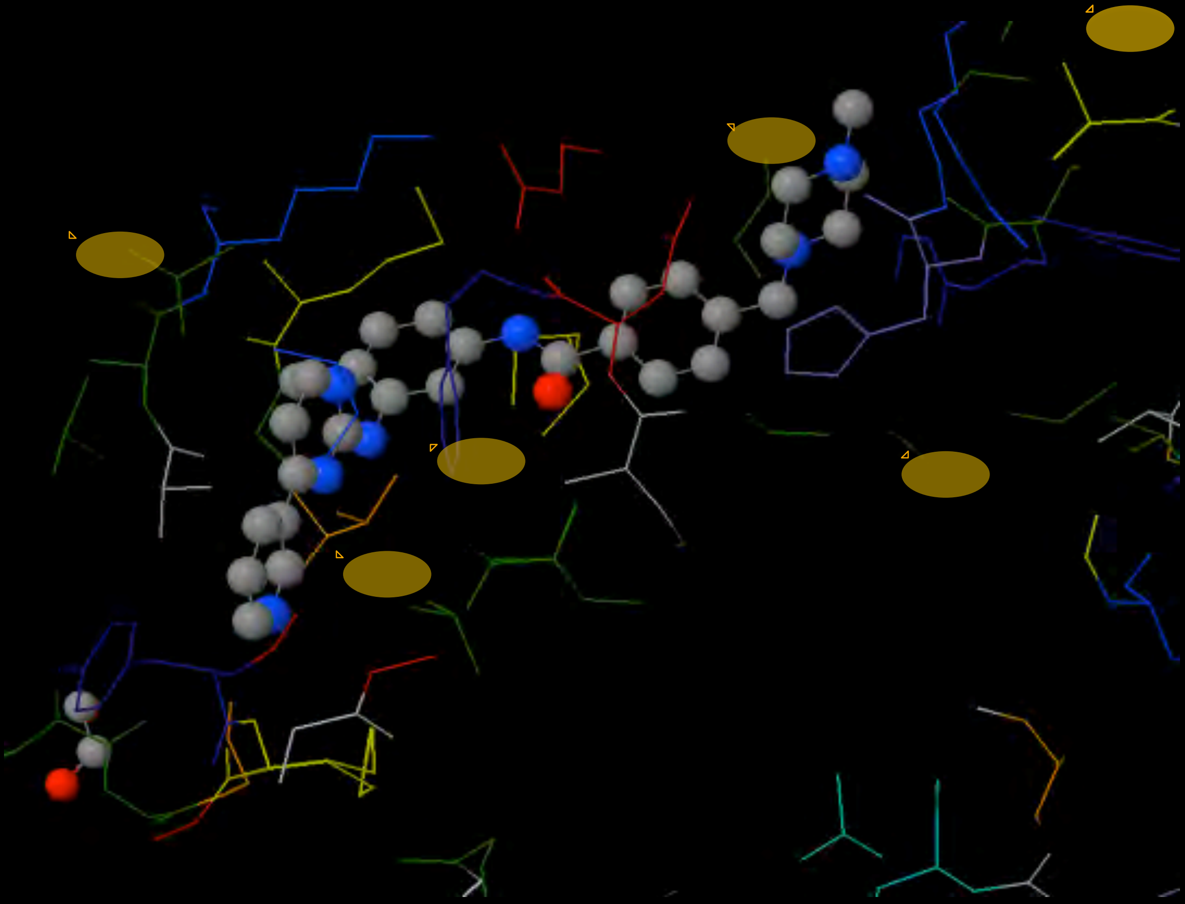
KiNEXUS

DrugKiNET KnowledgeBase

www.drugkinet.ca

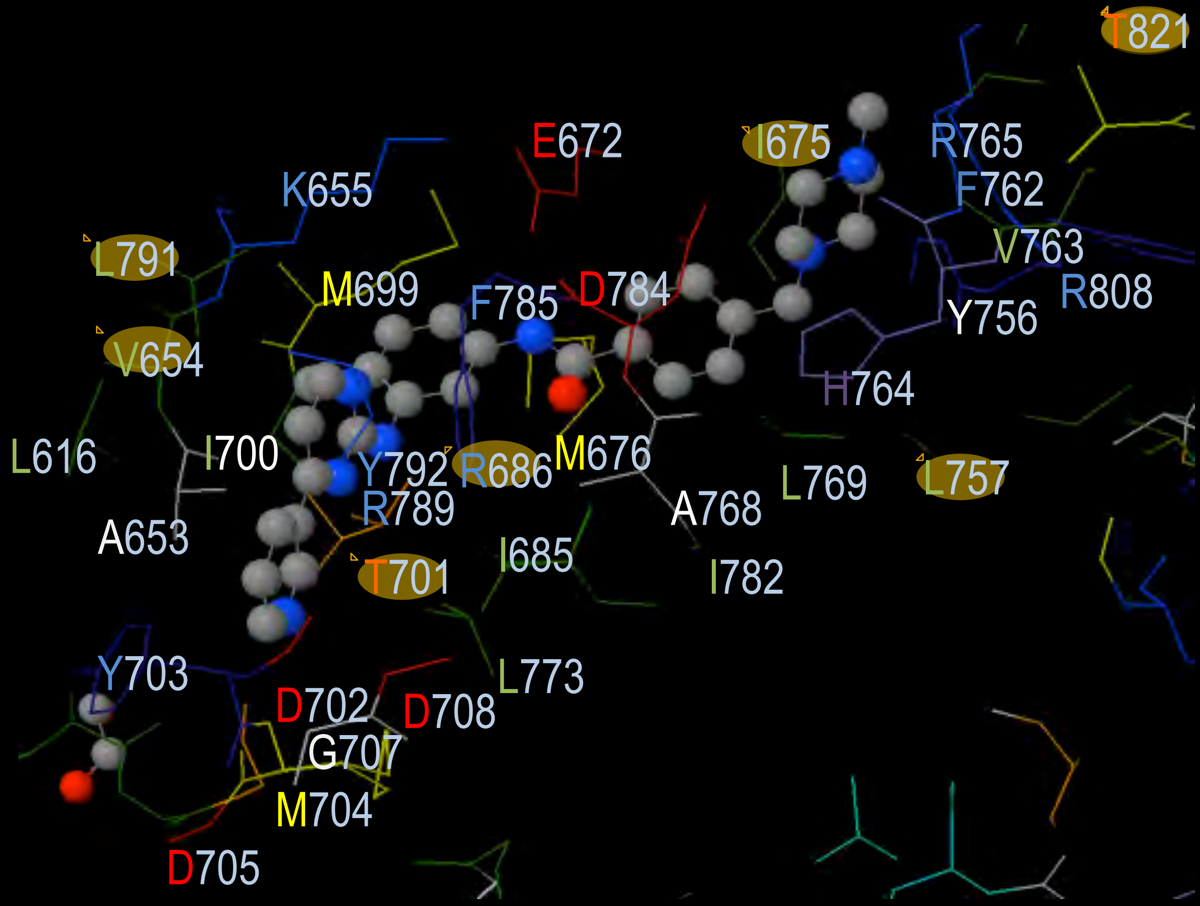
- Human protein kinase drug knowledgebase
- Features experimental results for over 105,000 kinase drug pairs (400 kinases, >850 compounds that inhibit kinases)
- Features predictive data for another 200,000 kinase-drug pairs with 500 kinases
- Information and links to other websites provided about each drug

Identified Imatinib Interactions with DDR1 Amino Acid Residues



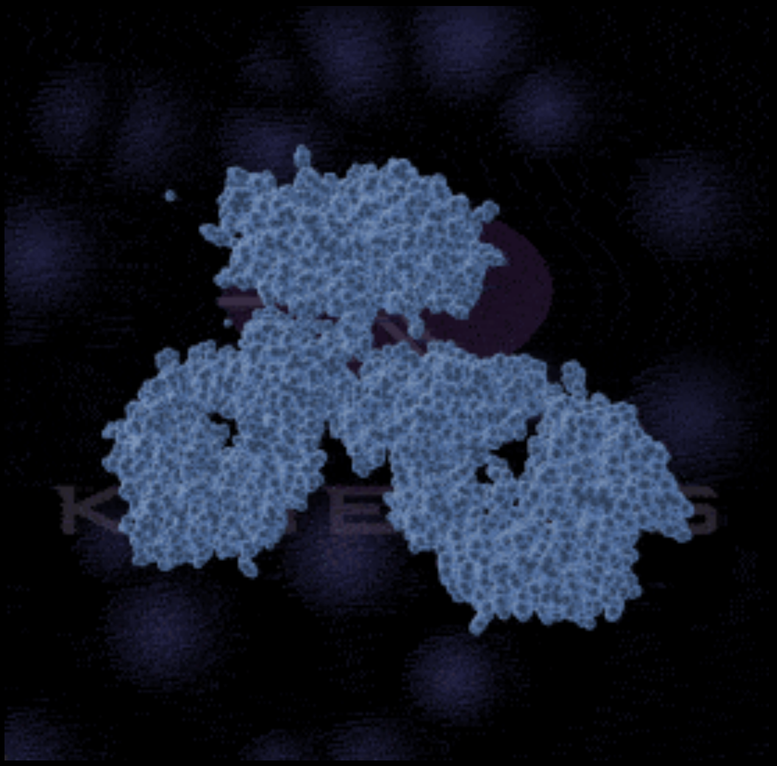
- DrugKiNET identified amino acids are indicated in orange ovals

Identified Imatinib Interactions with DDR1 Amino Acid Residues



- DrugKiNET identified amino acids are indicated in orange ovals
- Similar analyses may define biomarkers for drug sensitivity

Kinexus Antibody-based Analyses



- Kinetworks™ and Kinex™ proprietary analyses uses over 850 of the world's best antibodies
  - Targets low abundance cell signalling proteins
  - Quantifies protein expression and phosphorylation
  - Kinetworks™ multi-immunoblotting is more accurate and quantitative than competing methods
  - Kinex™ antibody microarray is more comprehensive and sensitive than other methods

KiNET-IB DataBase

www.kinet.ca

- Open access Kinetworks™ multi-immunoblotting database
- Features results from >10,000 immunoblots that track protein expression and phosphorylations from data generated with over 300 antibodies
- Over 200,000 protein measurements
- Searchable by protein, treatment or cell/tissue type
- Over 95% unpublished data

KiNEXUS

KiNET-AM DataBase

www.kinet-am.ca

KiNET™

Antibody Microarray Database

AN ONLINE RESOURCE FROM KINEXUS BIOINFORMATICS AND KINETICA ONLINE

- Open access Kinex™ KAM antibody microarray database
- Features results from >2,500 microarrays that track protein expression and phosphorylations data generated with 650 to 850 antibodies
- Over 3 million protein measurements
- Searchable by protein, treatment or cell/tissue type
- Over 99% unpublished data

Two mapping initiatives are currently underway at Kinexus

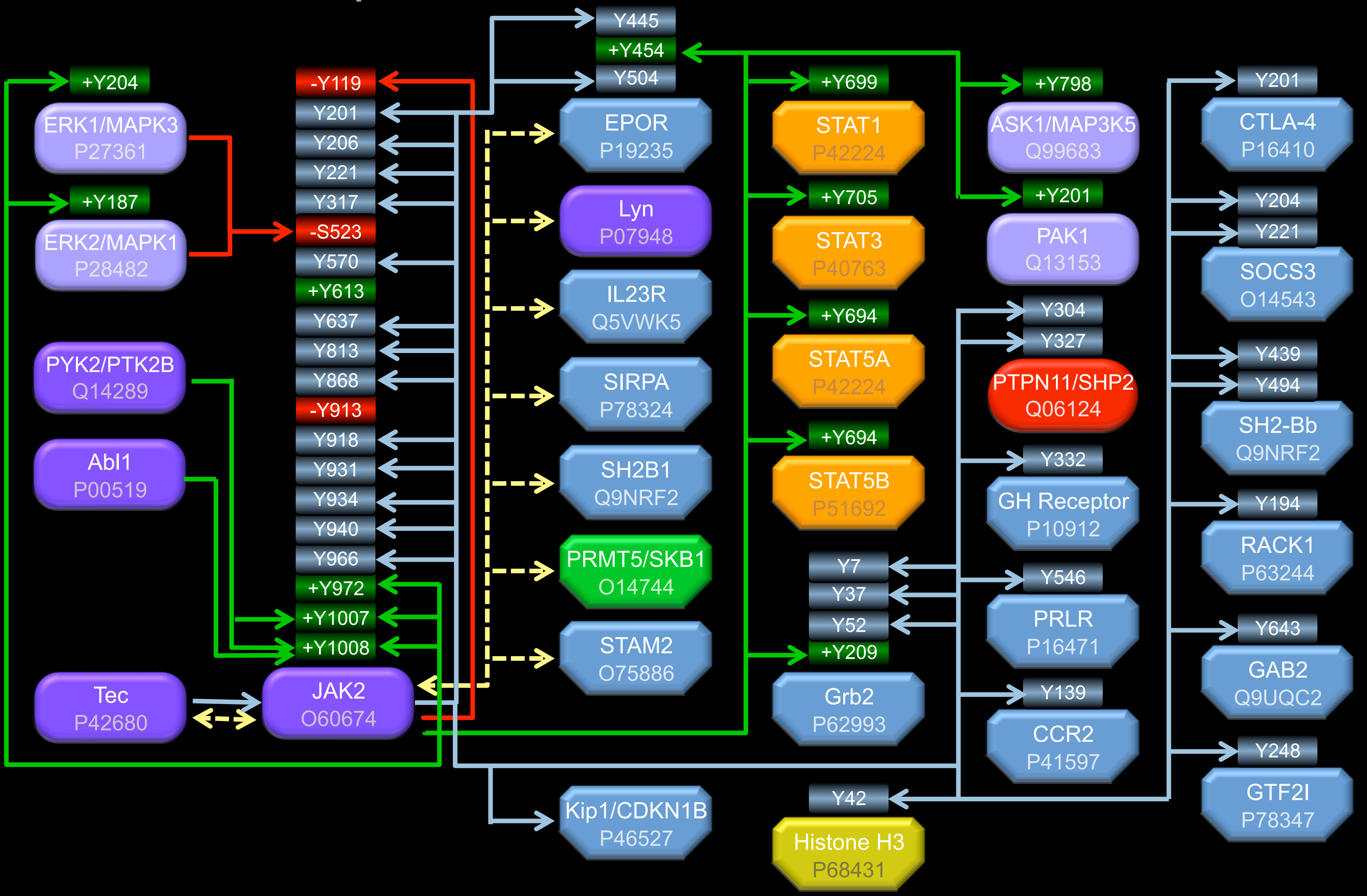
- Kinections Maps
  - Static, hand-annotated mini-maps that are target-centric with immediate upstream and downstream direct and functional interactions tracked
  - To be provided with Kinex™ KAM- 850 Antibody Microarray analyses – An MS-Excel data table is generated with each map

KINEXUS

Example of a Kinections Map

Kinections Map O60674

Janus Kinase 2



+Y204
ERK1/MAPK3
P27361
+Y187
ERK2/MAPK1
P28482
PYK2/PTK2B
Q14289
Abl1
P00519
Tec
P42680

-Y119
Y201
Y206
Y221
Y317
-S523
Y570
+Y613
Y637
Y813
Y868
-Y913
Y918
Y931
Y934
Y940
Y966
+Y972
+Y1007
+Y1008
JAK2
O60674

Y445
+Y454
Y504
EPOR
P19235
Lyn
P07948
IL23R
Q5VWK5
SIRPA
P78324
SH2B1
Q9NRF2
PRMT5/SKB1
O14744
STAM2
O75886
Kip1/CDKN1B
P46527

+Y699
STAT1
P42224
+Y705
STAT3
P40763
+Y694
STAT5A
P42224
+Y694
STAT5B
P51692
Y7
Y37
Y52
+Y209
Grb2
P62993
Y42
Histone H3
P68431

+Y798
ASK1/MAP3K5
Q99683
+Y201
PAK1
Q13153
Y304
Y327
PTPN11/SHP2
Q06124
Y332
GH Receptor
P10912
Y546
PRLR
P16471
Y139
CCR2
P41597

Y201
CTLA-4
P16410
Y204
Y221
SOCS3
O14543
Y439
Y494
SH2-Bb
Q9NRF2
Y194
RACK1
P63244
Y643
GAB2
Q9UQC2
Y248
GTF2I
P78347

- KinATLAS KnowledgeBase
  - Dynamic, customizable tissue/cell-specific maps of protein substrate and drug interactions with kinases based on experimental and predictive data
  - Identifies direct protein-protein interactions
  - Constructed from PhosphoNET, TranscriptoNET and DrugKiNET data and hyperlinks to these and other websites
  - Generates new maps with associated protein as the hub

KINEXUS

Example of KinATLAS Map

Select type of query desired

Query Type

Single Target Protein Phosphorylation Interaction

This query identifies the top phosphorylation interactions involving protein kinases and their substrates around a selected protein of interest. The target protein appears in the center of the generated interaction map that is cell, tissue or organ-specific. The stronger the kinase-substrate connection as calculated with our Kinase Substrate Predictor Version 2.0 Algorithm, the closer the associated protein. Apply the following filters to narrow down the most relevant interactions to your individual specifications.

Filters

Step 1 - Target Protein Identification

Provide the gene name, protein name, UniProt ID, RefSeq or IPI ID as a search term.

Step 2 - Human Cell/Tissue Specification

Use one of the two pull-down menus to specify the cell, tissue or organ type of interest.

Step 4 - Generation of Protein Phosphorylation Map

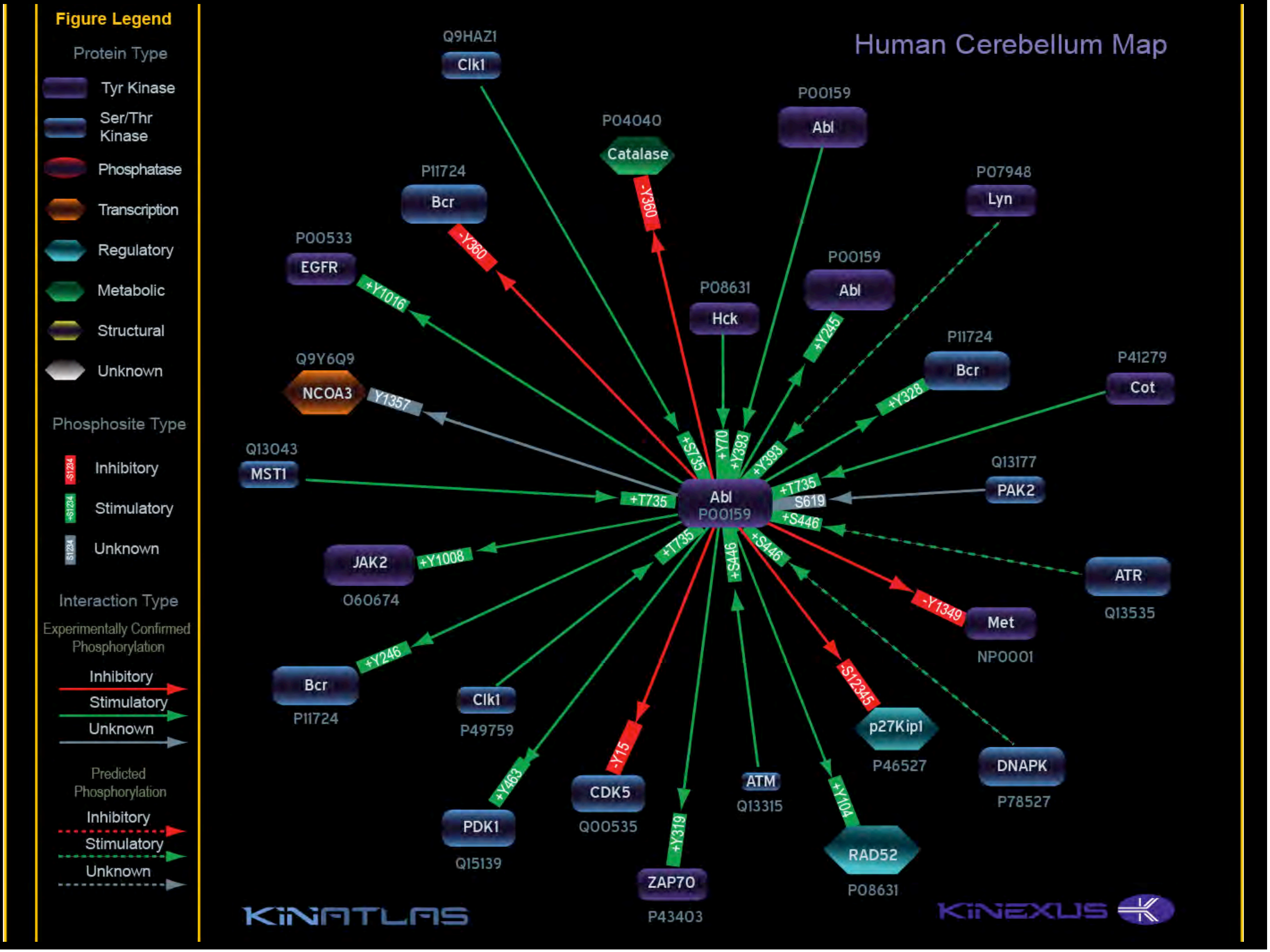
Click on boxes below to produce a custom map or reset parameters. The map will appear in a new window.

Step 3 - Selection of Display Parameters

Boxes that are marked are activated to display. Click on boxes to change status.

- ☑ Show experimentally-derived interactions
- ☑ Show predicted interactions
- ☑ Show upstream inputs
- ☑ Show downstream outputs
- ☑ Show stimulatory phosphosites
- ☑ Show inhibitory phosphosites
- ☑ Show non-defined phosphosites
- ☑ Show protein kinases
- ☑ Show protein phosphatases
- ☑ Show transcription factors
- ☑ Show regulatory proteins
- ☑ Show metabolic proteins
- ☑ Show structural proteins
- ☑ Show uncharacterized proteins
- ☑ Show black background
- ☐ Show icons with uniform size
- ☑ Show UniProt ID

Example of KinATLAS Map



Example of KinATLAS Map

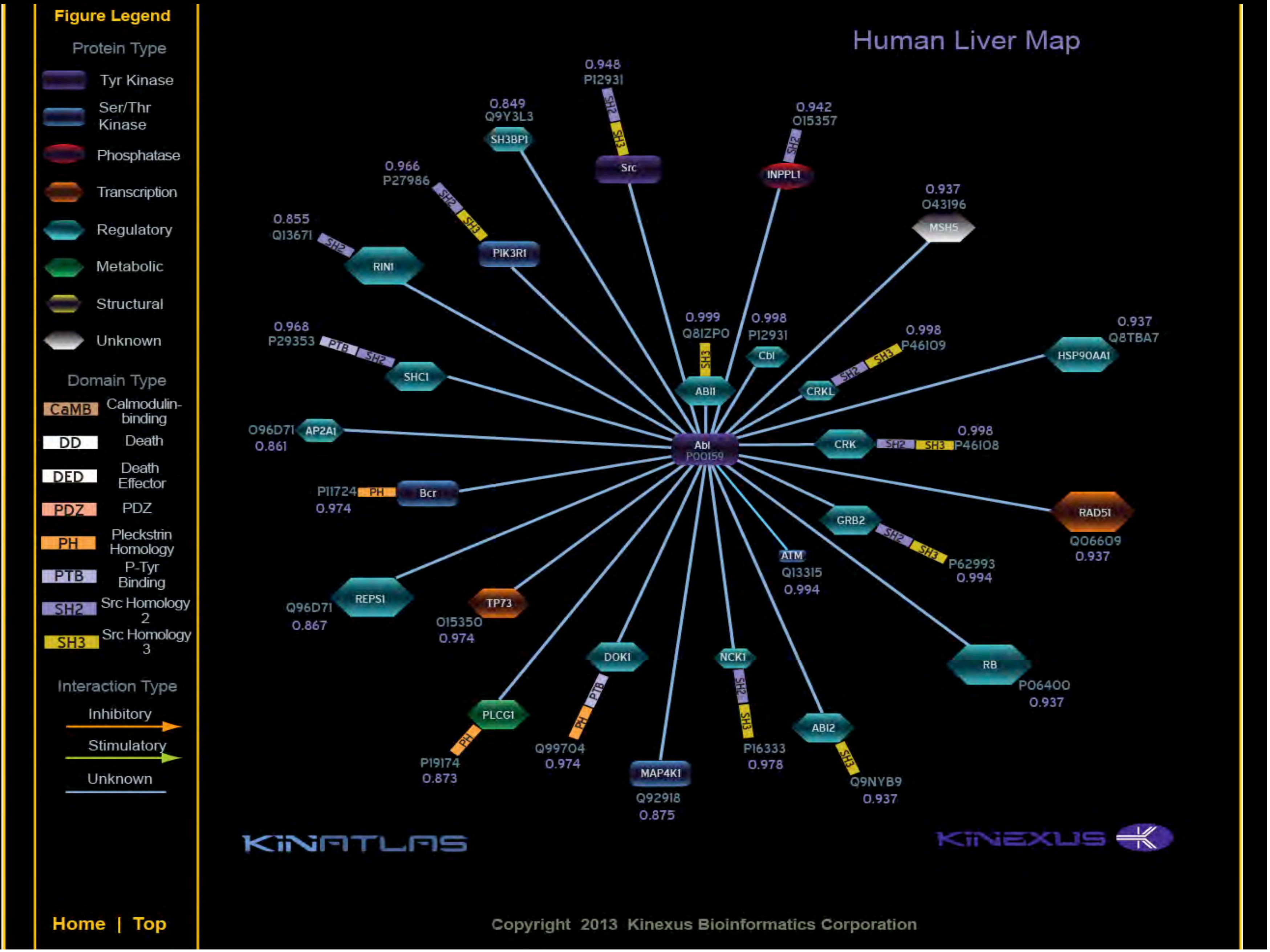


Figure Legend

Protein Type

- Tyr Kinase
- Ser/Thr Kinase
- Phosphatase
- Transcription
- Regulatory
- Metabolic
- Structural
- Unknown

Domain Type

CaMB	Calmodulin-binding
DD	Death
DED	Death Effector
PDZ	PDZ
PH	Pleckstrin Homology
PTB	P-Tyr Binding
SH2	Src Homology 2
SH3	Src Homology 3

Interaction Type

- Inhibitory
- Stimulatory
- Unknown

Home | Top

Example of KinATLAS Map

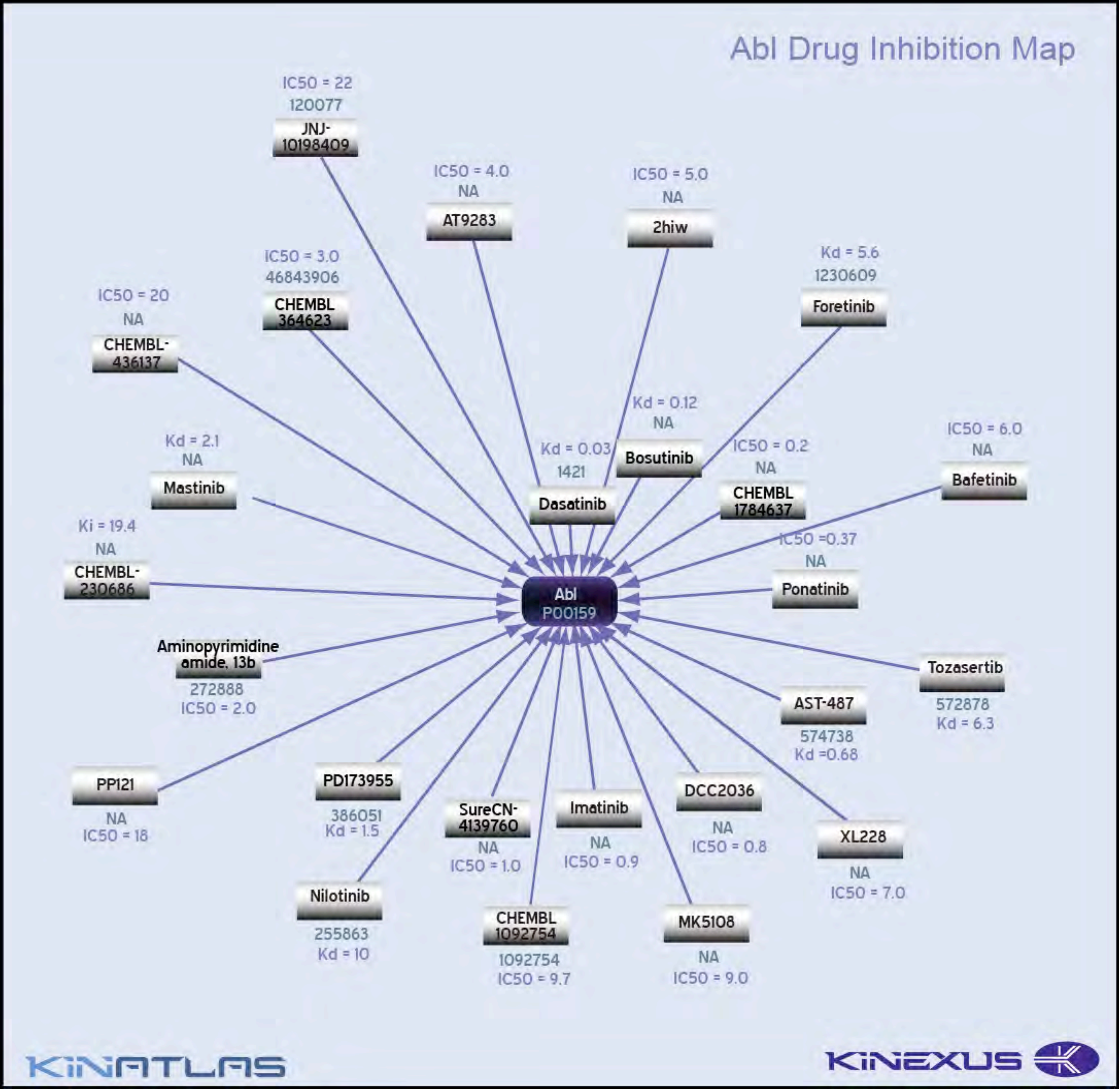
Figure Legend

Protein Type

- ▨ Tyr Kinase
- ▨ Ser/Thr Kinase

Interaction Type

- Experimentally Confirmed Inhibition →
- Predicted Inhibition ⇢



KiNEXUS

Credits

University of British Columbia

Javad Safaei, Dr. Jano Manuch, Shabab Hossein, Alireza Davoodi, Shenshen Lai, Sehyun Cho, Stephen Yiu, Chuen Choi, Pier-Luc Clemont, Christine Eisner, Cameron Rogers, Scott Wilder, Leon, Chew, Erandika Gunaretnam, Anna Isphording, Christopher Laver, Asher Mendelson, Steven Pham, Cecilia Sjoestroem, and Jovian Wat

Simon Fraser University

Dr. Ladislav Stacho

Kinexus Bioinformatics Corporation

Dr. Steven Pelech, Winnie So, Catherine Sutter, Dr. Hong Zhang, Dr. Dirk Winkler, Dr. Jane Shi, Dr. Arthur Yee, Dr. Rod Verdie, Cameron Bowyer, Litsa Blanis, Angie Chu, Mandy Chung, Pringle Comia, Aneesa Din, Nova Do, Bryanna Grace, Eileen Hu, Iris Juan, Joanna Kam, Stephanie Lam, Adam Leigh, Raymond Leung, Sara Marrello, Charrise Pagarigan, Andrew Park, Michael Peabody, Sabrina Rayworth, Ryan Saranchuk, Kevin Sha, Ellen Sung, Sharan Swarup, Amy Tam, Ryan Whitehead, Christine Yu, and Sarah Zaidi

British Columbia Institute of Technology

PhosphoNET (www.phosphonet.ca)

Joe Hu, Jonathan Jose, Ysabel Lago, David Liau, John Liau, Christine Livingstone, Melissa Manalac, Travis Nicholson, Kevin Odger, Ryan Pattinson, Nevin Petersen, Hannah Rosellon, Janice Sargent, Brandon Wong, and Alvin Yip

TranscriptoNET (www.transcriptonet.ca)

Isan Chen, Keegan Kelly, Aly Jamani, and Colin Nguyen

DrugKiNET (www.drugkinet.ca)

James Chen, Andy Chow, Sunju Christine Jeong, and Dan Stephenson

OncoNET (www.onconet.ca)

Kyle Li, Theo Mutia, Travis Ryder, and Clarence Sng

KiNET-AM (www.kinet-am.ca)

Simon Ho and the following team members: Nicholas Tagle, Fausto Faioli, Igor Kozlov, and Justin Ma