

Extraction of Functional Point Mutation Information from Biomedical Literature

Lawrence Lee
UCSF
Cohen Lab
NLM Conference 6/27/07

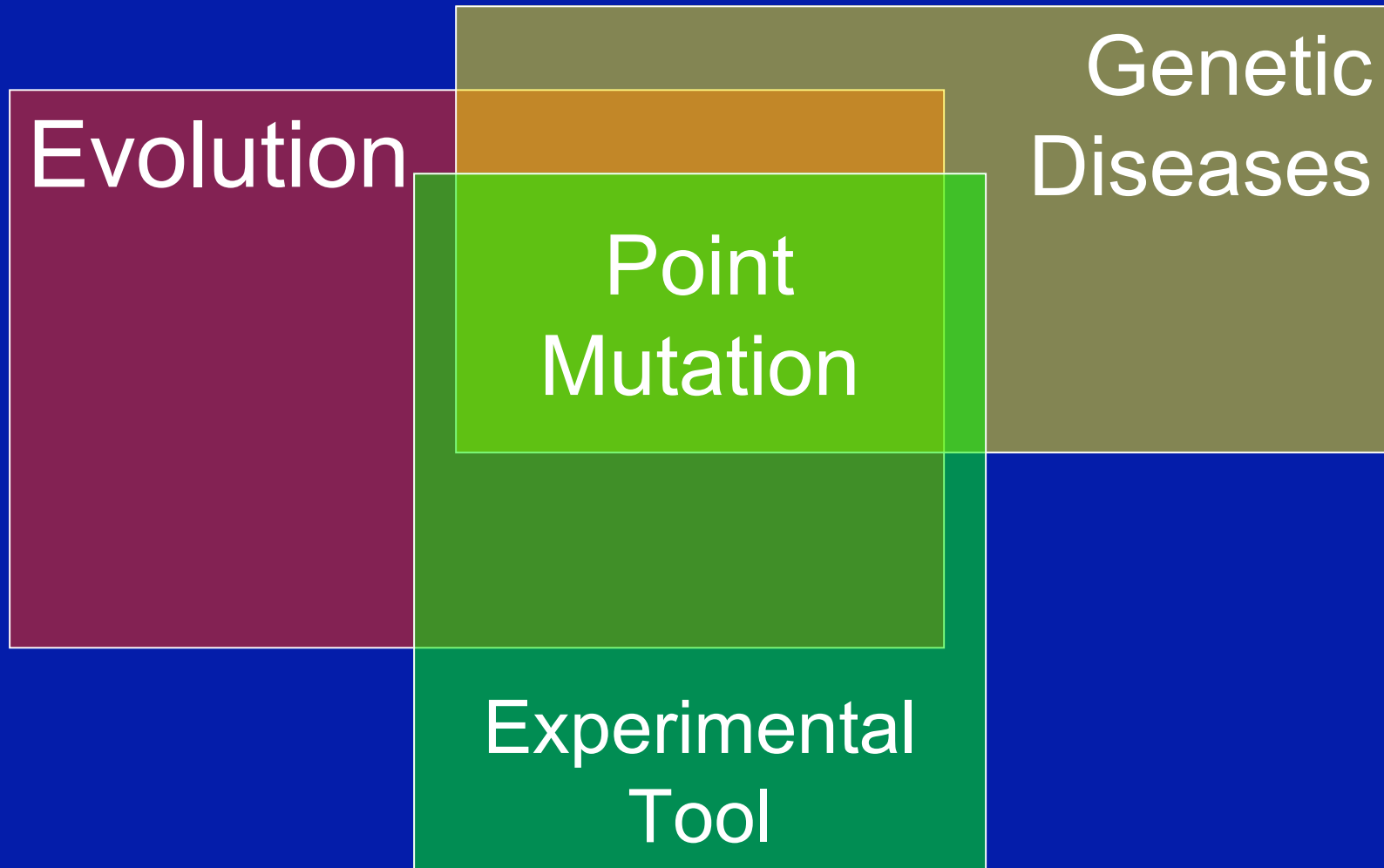
Importance of Point Mutations

Evolution

Genetic
Diseases

Point
Mutation

Experimental
Tool



Defined Structure



A(123)T	A123del	123insG	Ala123Thr
A123→T	ΔA123	DeltaA123	A123 to T

Identifying a Functional Motif and Its Effects

1

2

3

Mutations of the Conserved DRS Motif in the Second Intracellular Loop of the Gonadotropin-Releasing Hormone Receptor Affect Expression, Activation, and Internalization

Krishan K. Arora, Zhengyi Cheng, and Kevin J. Catt

Endocrinology and Reproduction Research Branch
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, Maryland 20892

The GnRH receptor is an unusual member of the G protein-coupled receptor (GPCR) superfamily with several unique features. One of these is a variant of the conserved DRY motif that is located at the junction of the third transmembrane domain and the second intracellular (2i) loop of most GPCRs. In the GnRH receptor, the Tyr residue of the conserved triplet is replaced by a DRS sequence. The aspartate (Asp) at D138N is highly conserved in all GPCRs. The functional importance of these residues was evaluated in wild-type and mutant GnRH receptors expressed in *Xenopus* oocytes. Replacing Asp138 with Ala or Ser showed reduced internalization, and the GnRH-induced inositol phosphate response for the Arg138Gln mutant was significantly impaired in proportion to its low expression level. Replacing Ser151 with Ala or Glu showed reduced internalization and signal transduction. In two additional mutants, Ser151Ala and Ser151Glu, Ser151Ala, Ser151Glu, and Glu156Gln. Both of these mutants showed reduced internalization and signal transduction. The conserved Asp and Arg residues in the DRY/S triplet make important contributions to the structural integrity of the receptor and influence its function.

receptor expression, agonist-induced activation, and internalization. (Molecular Endocrinology 11: 1203-1212, 1997)

INTRODUCTION

The hypothalamic decapeptide, GnRH, acts via its specific high-affinity receptors in the anterior pituitary gland to regulate the synthesis and secretion of FSH and LH and thus plays a pivotal role in reproduction (1, 2). The cloning of cDNAs for the GnRH receptors of several species, including mouse (3, 4), rat (5-7), sheep (8, 9), cow (10), and human (11, 12), has shown that the receptor exhibits more than 85% amino acid identity among species. The hydrophathy analysis of the GnRH receptor reveals the presence of seven transmembrane domains (TM I-VII), indicating a similar topology to those proposed for the other members of the G protein-coupled receptor (GPCR) superfamily (13). However, the GnRH receptor has several unique features, including the absence of a cytoplasmic carboxyl-terminal tail, replacement of Tyr by Ser in the highly conserved DRY sequence located at the junction of TM III and the 2i loop, and the presence of a long and highly basic first intracellular loop. Another interesting feature is that the highly conserved Asp in TM II and Asn in TM VII of most GPCRs are reciprocally exchanged in the GnRH receptor (13, 14). Mutagenesis and chimeric studies have suggested that the intracellular regions of the GPCRs, in particular the second and third intracellular (2i and 3i) loops and sometimes the cytoplasmic tail, interact with G proteins and mediate signal transduction (15-18). Sequence alignment of various members of the GPCR superfamily shows that the acidic (Asp) and basic (Arg) residues of the DRY triplet are highly conserved (15,

0896-8959/97/\$3.00/0
Molecular Endocrinology
Copyright © 1997 by The Endocrine Society

Mutations of the Conserved DRS Motif in the Second Intracellular Loop of the Gonadotropin-Releasing Hormone Receptor Affect Expression, Activation, and Internalization

Krishan K. Arora, Zhengyi Cheng, and Kevin J. Catt

Endocrinology and Reproduction Research Branch
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, Maryland 20892

Mutants in which Asp138 was replaced by Asn or Glu were poorly expressed, but showed significantly increased internalization and exhibited augmented inositol phosphate generation to maximal agonist stimulation compared with the wild type receptor.

Replacing Ser140 with Ala affected neither internalization nor signal transduction.

Replacing Ser140 with Ala affected neither internalization nor signal transduction.

Replacing Ser140 with Ala affected neither internalization nor signal transduction.

0896-8959/97/\$3.00/0
Molecular Endocrinology
Copyright © 1997 by The Endocrine Society

Mutations of the Conserved DRS Motif in the Second Intracellular Loop of the Gonadotropin-Releasing Hormone Receptor Affect Expression, Activation, and Internalization

Mutation	Verb	Noun Phrase
D138N	show	significantly increased internalization
D138N	exhibit	augmented inositol phosphate generation
S140A	↯affect	internalization
S140A	↯affect	signal transduction

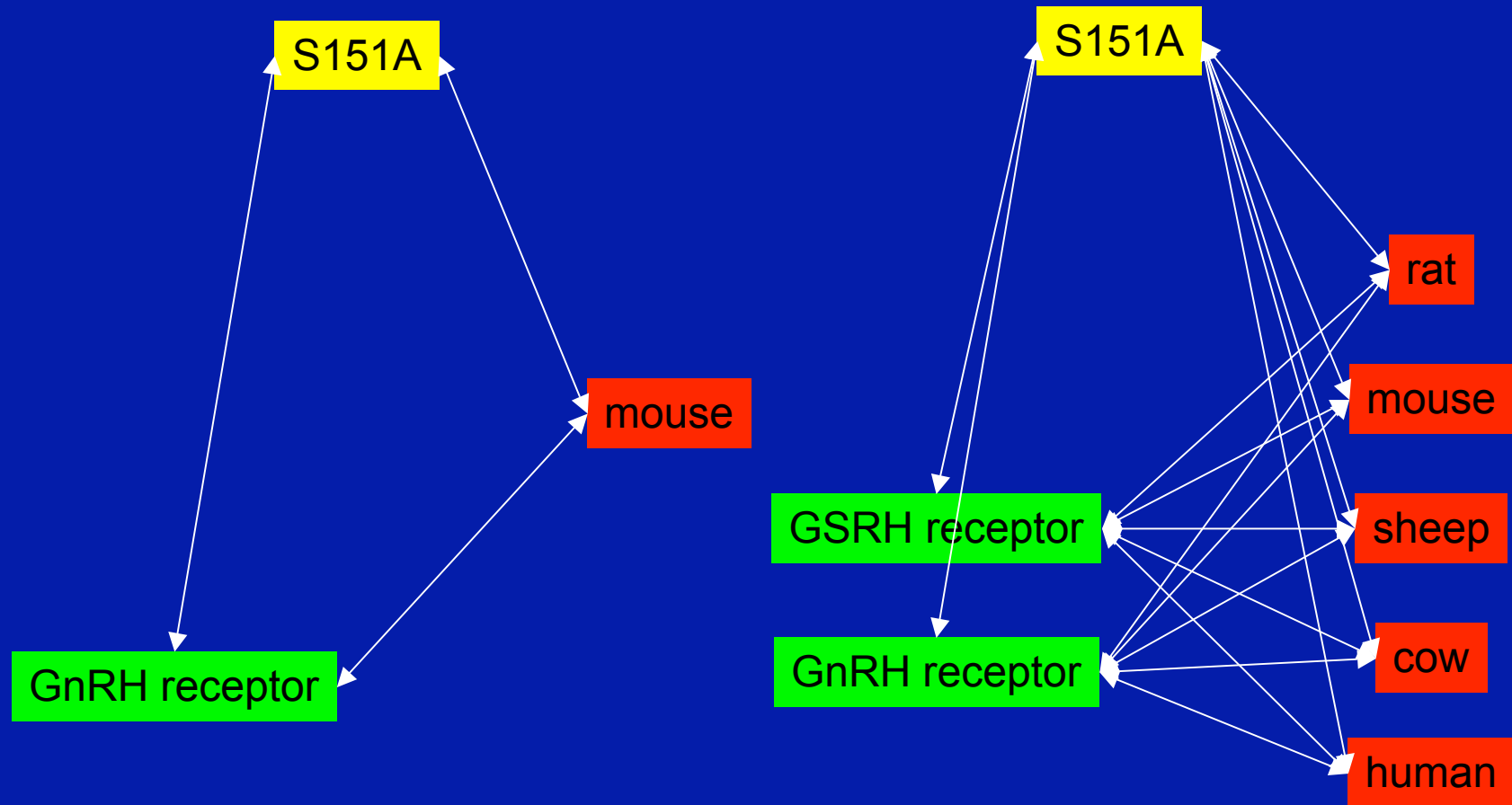
0896-8959/97/\$3.00/0
Molecular Endocrinology
Copyright © 1997 by The Endocrine Society

S151A → Q01776

1

Identifying Mutation → Protein Pairs

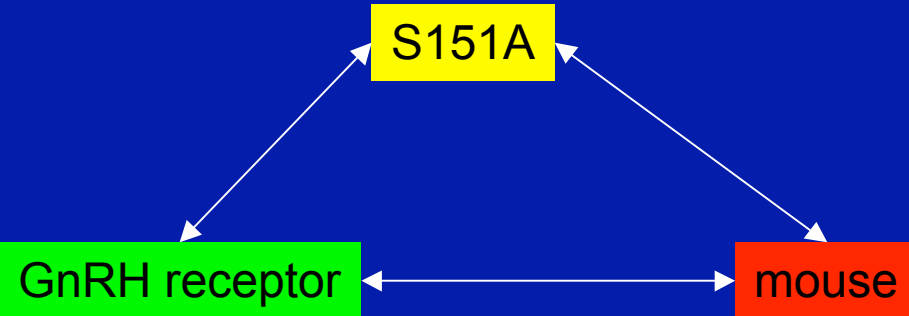
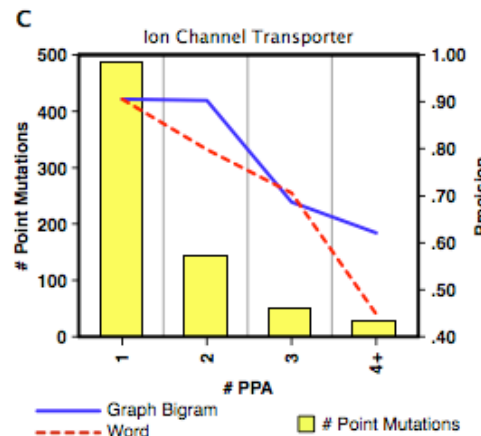
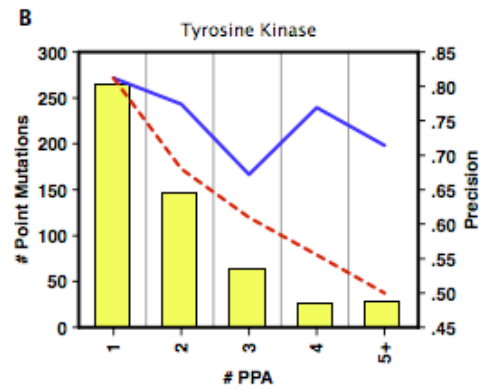
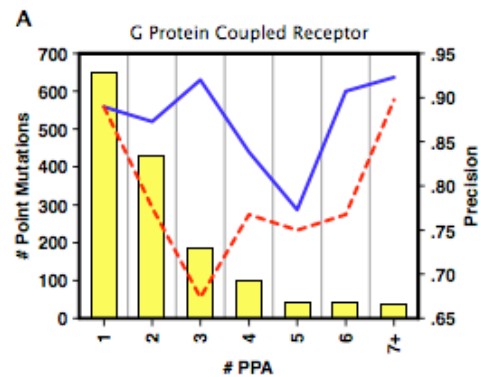
Protein Disambiguation



Q01776

P30969, Q01776, P32237,
P32236, Q92847

Graph-Bigram Search



$$h_{graph}(w_i, w_j) = \left[\frac{\tilde{x} - \mu}{\sqrt{s^2/N}} \right]^{-1}$$

Lee LC, Horn F, Cohen FE (2007) Automatic Extraction of Protein Point Mutations Using a Graph Bigram Association. PLoS Comput Biol 3(2): e16 doi:10.1371/journal.pcbi.0030016

2

Identifying Functional Effect Containing Sentences

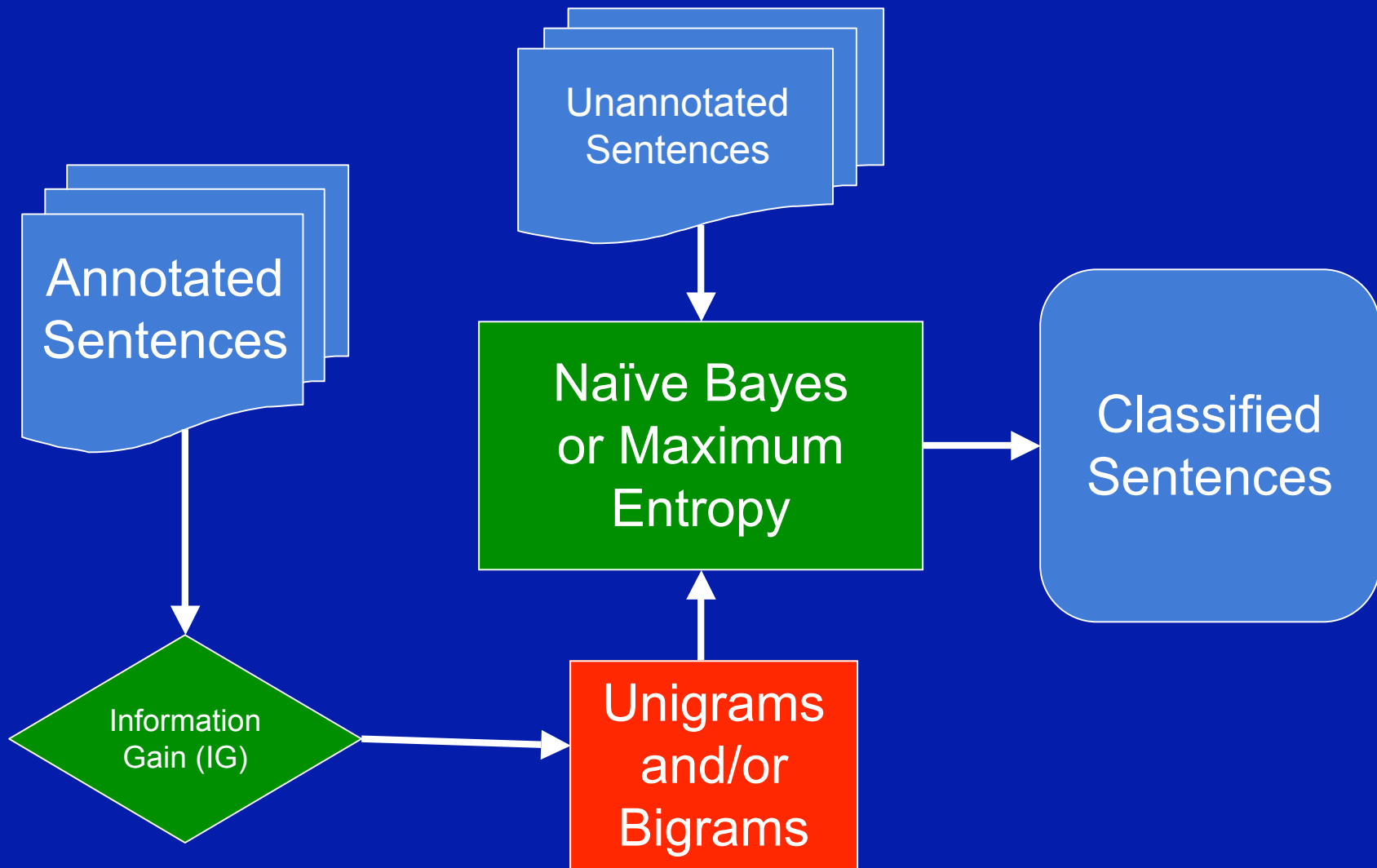
Functional Description?

- What can be classified as a mutation functional effect?
 - Qualitative or quantitative change in the biochemical, structural, enzymatic, and/or physical properties of a protein.
 - Change in properties of host cell, virus, tissue, and/or organ.
 - Change in interactions of protein with other members of its environment.

Non-functional?

- Population genetics (prevalence)
- Patient screening

Functional Classification



Sentence Datasets

PubMed Abstracts	Training Set	Test Set
“Functional”	477	538
“Non-functional”	522	476
Total	999	1014

CF Full Text Articles	Training Set	Test Set
“Functional”	192	168
“Non-functional”	372	302
Total	564	470

Information Gain (IG)

Change in entropy of a system when a variable is given.

$$\text{IG}(\text{Class}, \text{Attribute}) = H(\text{Class}) - H(\text{Class} | \text{Attribute})$$



that	method
increase	mutate
A337T	inhibit
activate	stability
bind	F508Del
reduce	CF patient
cfr	K1250A
patient	Ca+
exon	class IV

Abstract Sentences Naïve Bayes Classifier

	Unigram	Unigram + Bigram
Precision	0.836	0.849
Recall	0.814	0.805
F-measure	0.825	0.826

Abstract Sentences Maximum Entropy Classifier

	Unigram	Unigram + Bigram
Precision	0.862	0.883
Recall	0.763	0.730
F-measure	0.809	0.800

CFTR Sentences Naïve Bayes Classifier

	Unigram	Unigram + Bigram
Precision	0.789	0.859
Recall	0.887	0.909
F-measure	0.835	0.880

CFTR vs. Abstract Sentences Naïve Bayes

Unigram / Bigram		Trained On	
		Abstract Training Set	CFTR Training Set
Tested On	Abstract Test Set		0.709 / 0.664
	CFTR Test Set	0.694 / 0.692	

Errors/Improvements

- False positive/negative sentences tend to be very short or very long.
- Semantic information not included.

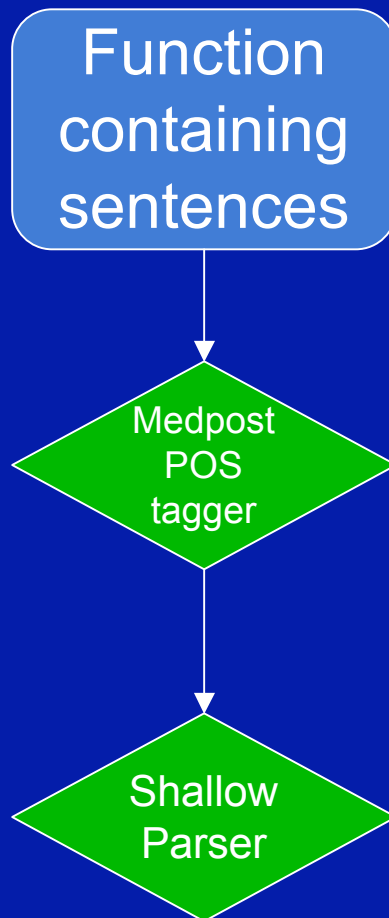
3

Extracting Functional Effects of Point Mutations

Desired Information

PMID	AC	Mutation	Verb	Noun Phrase
9259312	Q01776	D138N	show	significantly increased internalization
9259312	Q01776	D138N	exhibit	augmented inositol phosphate generation
9259312	Q01776	S140A	¬affect	internalization
9259312	Q01776	S140A	¬affect	signal transduction

Methods



We chose to study the R117H CFTR, which is known to localize to the cell surface and maintain normal PKA-dependent activation but which has reduced single-channel Cl conductance.

('We', 'PRP') ('chose', 'VBP') ('to', 'TO') ('study', 'VB') ('the', 'DT') ('R117H', 'NN') ('CFTR', 'NN') (',', ',') ('which', 'WDT') ('is', 'VBZ') ('known', 'VBN') ('to', 'TO') ('localize', 'VB') ('to', 'TO') ('the', 'DT') ('cell', 'NN') ('surface', 'NN') ('and', 'CC') ('maintain', 'VBP') ('normal', 'JJ') ('PKA-dependent', 'JJ') ('activation', 'NN') ('but', 'CC') ('which', 'WDT') ('has', 'VBZ') ('reduced', 'JJ') ('single-channel', 'NN') ('Cl', 'NN') ('conductance', 'NN') (',', ',')

Methods (cont.)

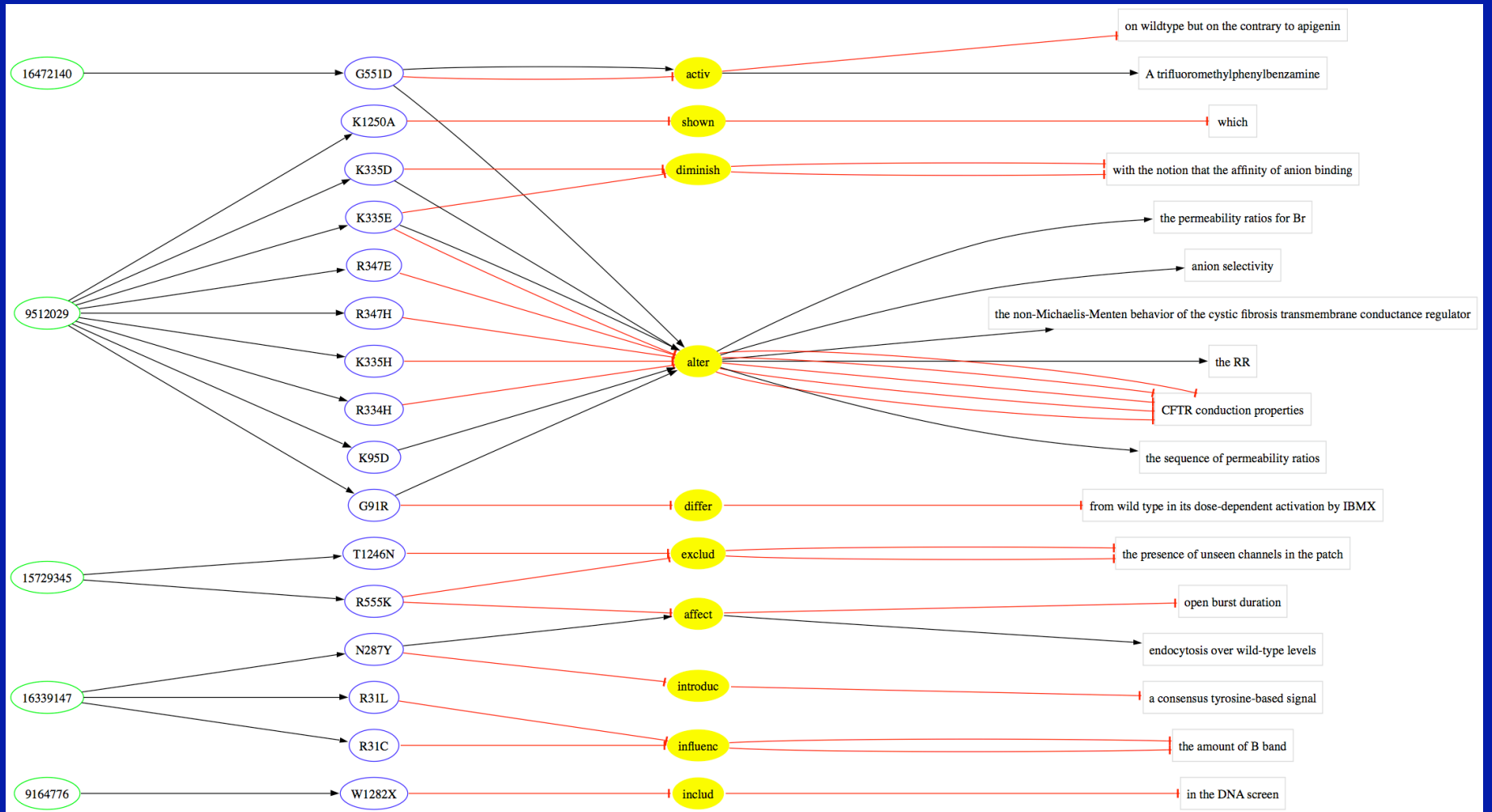
(S:
 (NP: ('We', 'PRP'))
 (VP: ('chose', 'VBP') ('to', 'TO') ('study', 'VB'))
 (MP: ('the', 'DT') ('R117H', 'NN') ('CFTR', 'NN'))
 (',', ','))
 (NP: ('which', 'WDT'))
 (VP: ('is', 'VBZ') ('known', 'VBN') ('to', 'TO')
 ('localize', 'VB'))
 (NP:
 (PP: ('to', 'TO'))
 (NP: ('the', 'DT') ('cell', 'NN') ('surface', 'NN'))))
 ('and', 'CC')
 (VP: ('maintain', 'VBP'))
 (NP: ('normal', 'JJ') ('PKA-dependent', 'JJ')
 ('activation', 'NN')) ('but', 'CC')
 (NP: ('which', 'WDT'))
 (VP: ('has', 'VBZ'))
 (NP: ('reduced', 'JJ') ('single-channel', 'NN')
 ('Cl', 'NN') ('conductance', 'NN'))
 (',', ','))

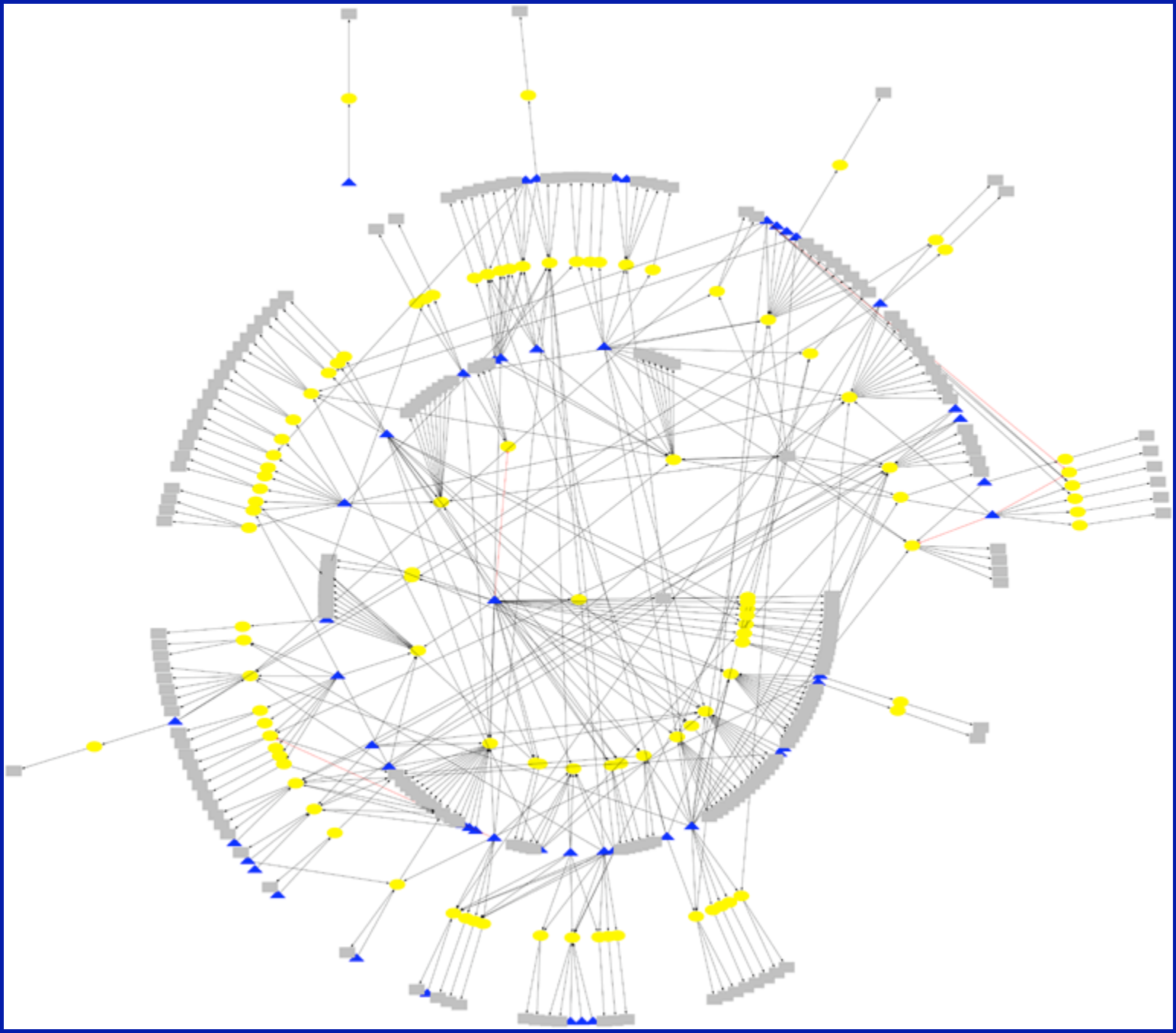
Pattern extraction:

<MP><VP><NP>

Mutation	Verb	Noun Phrase
R117H	localize	to the cell surface
R117H	maintain	normal PKA-dependent activation
R117H	has	reduced single-channel Cl conductance

“Functional” Network





Acknowledgements

Cohen Lab

Barney May

Jay Choi

Cedric Govaerts

Florence Horn

Jenni Weisman

Fred Cohen

NIH/NLM Funding

Long Sentence

- We performed a case-control study to test the association between two polymorphisms in the hMSH2 gene: an A --> G transition at 127 position producing an Asn --> Ser substitution at codon 127 (the Asn127Ser polymorphism) and a G --> A transition at 1032 position resulting in a Gly --> Asp change at codon 322 (the Gly322Asp polymorphism) and breast cancer risk and cancer progression.

Short Sentences

- HBV with TCC at nucleotides 1856 to 1858 was associated with the G1898A mutation (64%).
- Conversely, the BFD S26A variant was unable to cleave benzoin.
- RafS621A downregulated ERalpha and ERbeta expression.