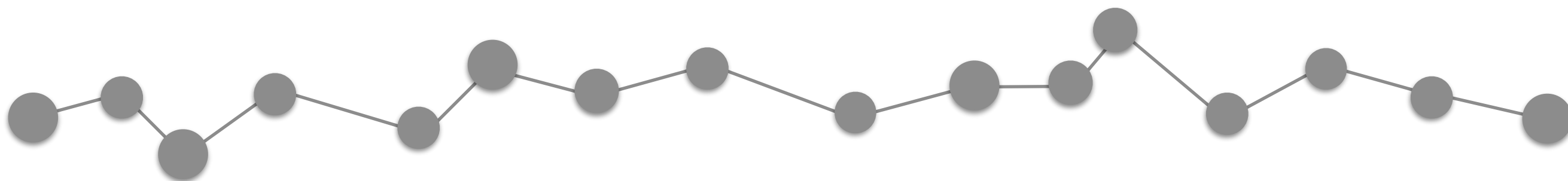


Hunting Single Point Mutations

PIUG BIOTECH 2016

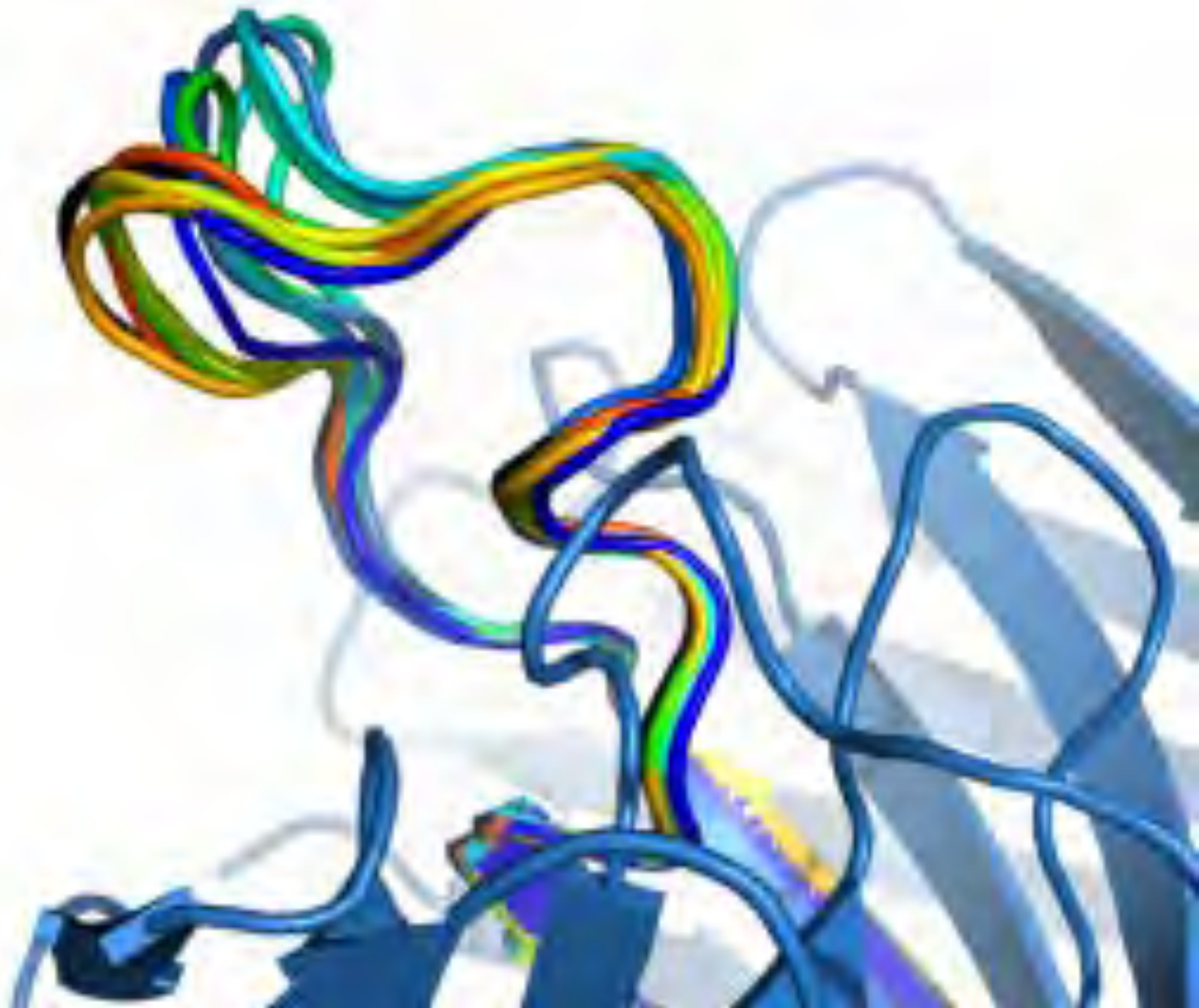
Julien Doudement

Why?



It's a common use case and it's not trivial



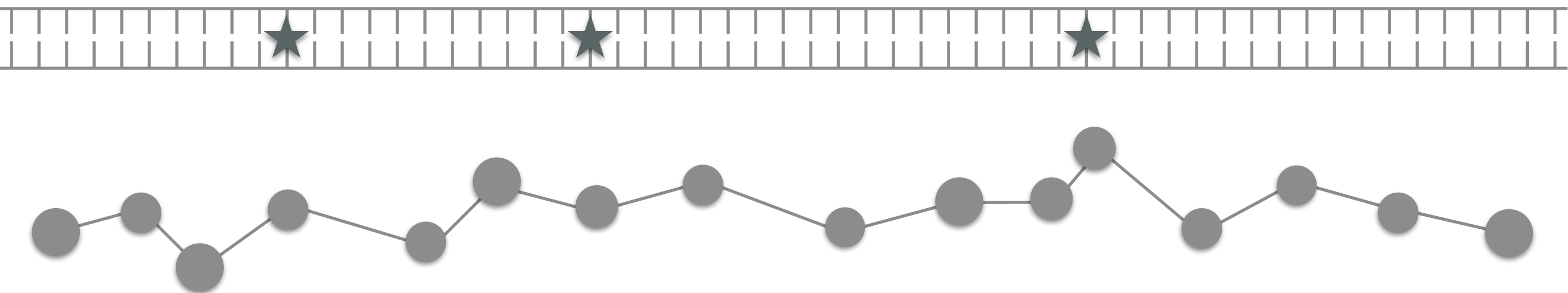


Antibody - CDRs



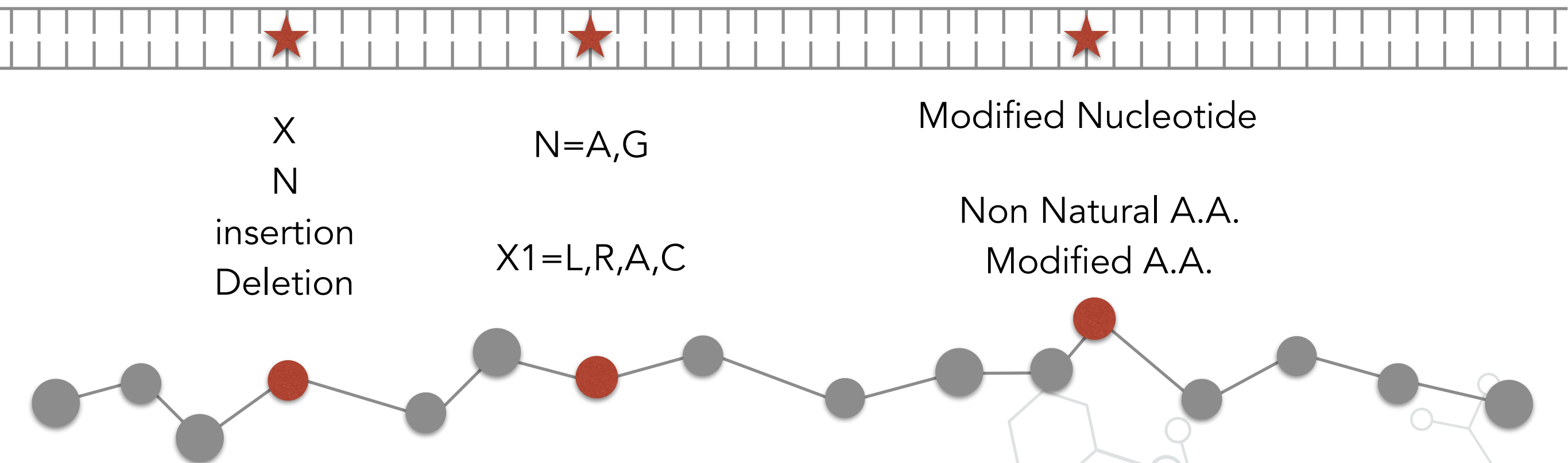
CRISPR / Cas9

Why?



The need to find sequences with specific variations at specific positions won't go away

Complexity (1)



What kind of variation/mutation are you chasing?

The Impact of WIPO Standards on Sequence Searching

for
PIUG Biotechnology Meeting 2014
Presented by Tara Rix & Tasha Witham

The current nomenclature InterpretationS

WIPO ST.25
Original Sequence

X_1 NRGNRKL X_2 NLRP X_3

X_1 = Norleucine

X_2 = R, L, P or N

X_3 = Norleucine, R, L, P or N

Interpretation comparison

HIC
Harbor Consulting IP Services, Inc.
The Sequence Listing and Search Experts

LNRGNRKL X_2 NLRP X_3

USPTO

X_1 NRGNRKLRLRPR

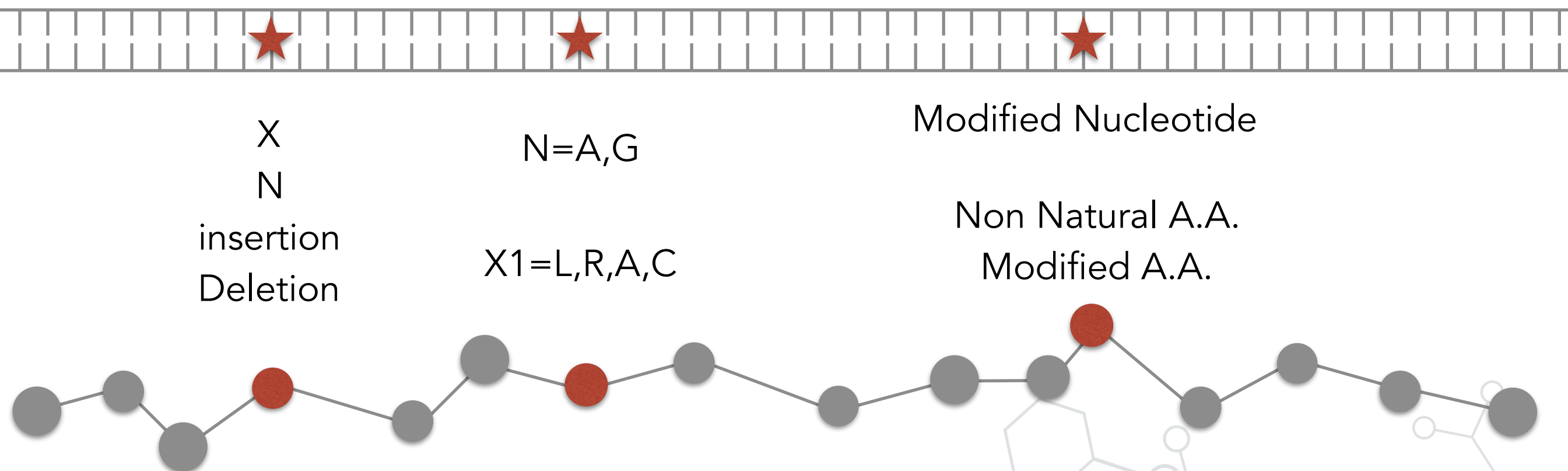
EPO

X_1 NRGNRKL X_2 NLRP X_3

PCT

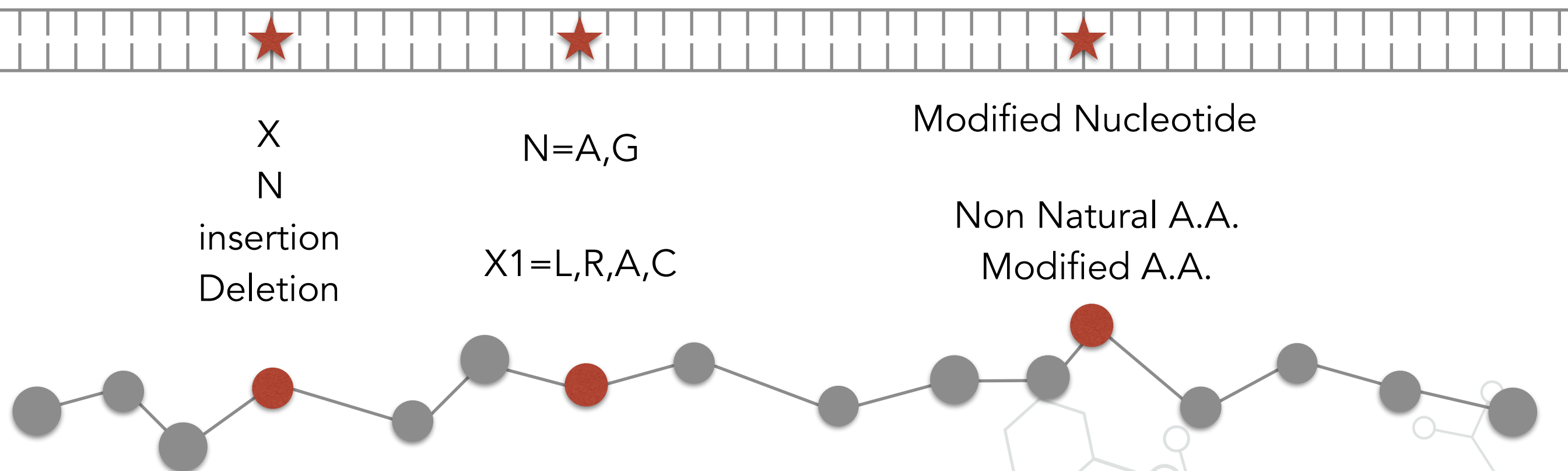
*Now, take a look at the
differences between
these sequences:*

Complexity (1)



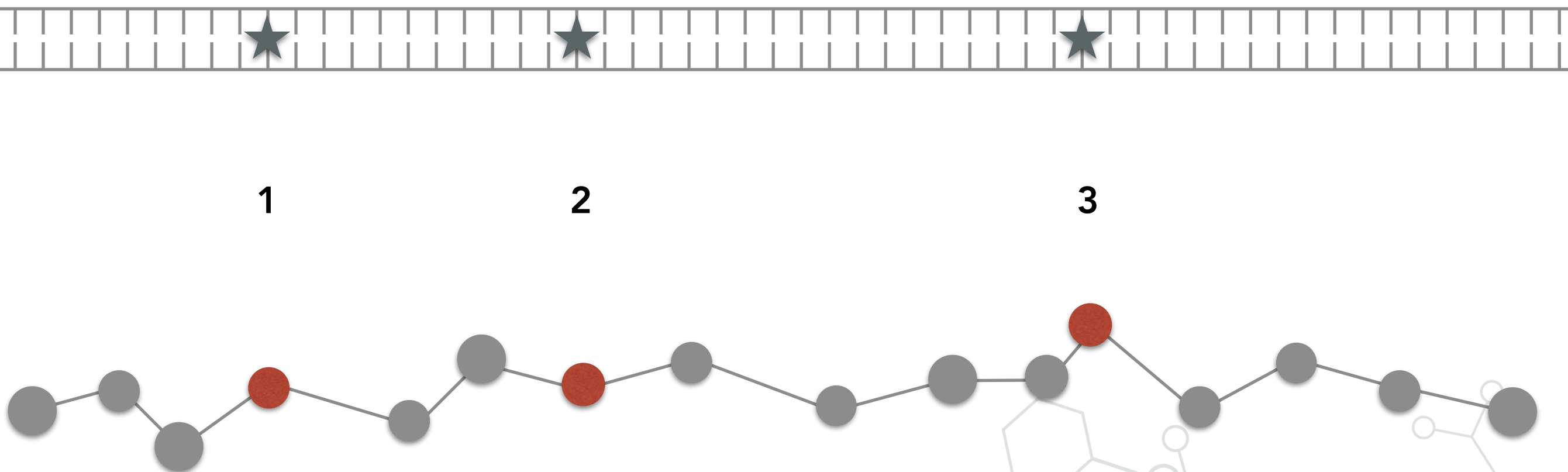
Always handle any of them as a mismatch.
Lower the homology accordingly.

Complexity (1)



You will always need to read the document to find the very details

Complexity (2)



Are you looking at combinations?

1 and 2 and 3?

1 and/or 2 and/or 3?

Complexity (3)



?

1

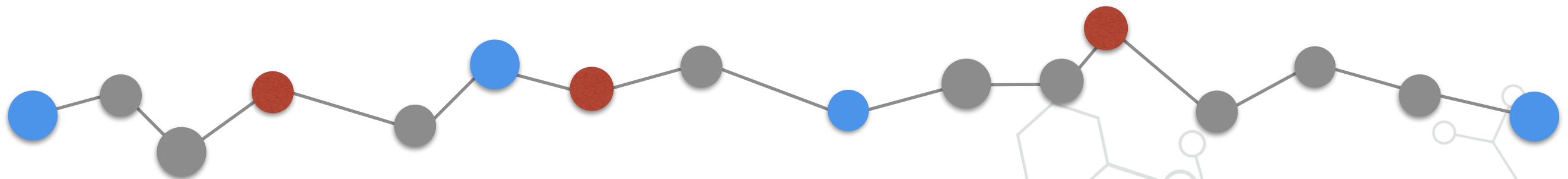
?

2

?

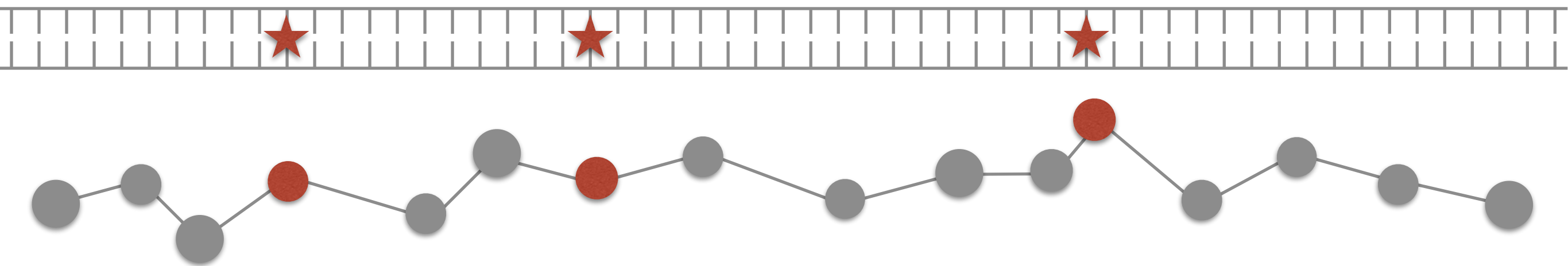
3

?



Would you accept extra variations?

Solutions (tricks?)



A set of 6 mutations can convert SpCas9 into a nicking enzyme

(Known information used to illustrate a sequence search)

>WT_CRISPR-associated endonuclease Cas9

```
MDKKYSIGL D IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA
LLFDSGETAE ATRLKRTARR RYTRRKNRIC YLQEIFSDEM AKVDDSFHHR 100
LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD
LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP 200
INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP
NFKSNFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI 300
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI
FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR 400
KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY
YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASASQ FIERMTNFDK 500
NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD
LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI 600
IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ
LKRRRYTGWG RLSRKLINGI RDKQSGKITL DFLKSDGFAN RNFMQLIHDD 700
SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVDELVKV
MGRHKPENIV IEMARENQTT QKGQKNSRER MKRIEEGIKE LGSQILKEHP 800
VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVDH IVPQSFLKDD
SIDNKVLTRS DKNRGKSDNV PSEEVVKMK NYWRQLLNAK LITQRKFDNL 900
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI
REVKVITLKS KLVSDFRKDF QFYKVVREINN YHHAH DAYLN AVVG TALIKK 1000
YPKLESEFVY GDYKVYDVRK MIAKSEQEIG KATAKYFFYS NIMNFFKTEI
TLANGEIRKR PLIETNGETG EIVWDKGRDF ATVRKVLSMP QVNIVKKTEV 1100
QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPTVA YSVLVVAKVE
KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDIIKLPK 1200
YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE
DNEQKQLFVE QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHARDK 1300
PIREQAENII HLFTLTNLGA PAAFKYFDTT IDRKYRSTK EVLDATLIHQ
SITGLYETRI DLSQLGGD
```

D10A
E762A
H840A
N854A
N863A
D986

Has this combination
or any of those
variations been
described already?

A set of 6 mutations can convert SpCas9 into a nicking enzyme

(Known information used to illustrate a sequence search)

>WT_CRISPR-associated endonuclease Cas9

MDKKYSIGL	IGTNSVGWAV	ITDEYKVPSK	KFKVLGNTDR		
LLFDSGETAE	ATRLKRTARR	RYTRRKNRIC	YLQEIFSNE	D10A	100
LEESFLVEED	KKHERHPIFG	NIVDEVAYHE	KYPTIYHLR	E762A	200
LRLIYLALAH	MIKFRGHFLI	EGDLNPDNSD	VDKLFQILV	H840A	300
INASGVDAKA	ILSARLSKSR	RLENLIAQLP	GEKKNGLFG	N854A	400
NFKSNFDLAE	DAKLQLSKDT	YDDDLNLLA	QIGDQYADLF	N863A	500
LLSDILRVNT	EITKAPLSAS	MIKRYDEHHQ	DLTLLKALVR	D986	600
FFDQSKNGYA	GYIDGGASQE	EFYKFIKPIL	EKMDGTEELL		
KQRTFDNGSI	PHQIHLGELH	AILRRQEDFY	PFLKDNREKI		
YVGPLARGNS	RFAWMTRKSE	ETITPWNFEE	VVDKGASASQ		
NLPNEKVLPK	HSLLYEYFTV	YNELTKVKYV	TEGMRKPAFL		
LLFKTNRKVT	VKQLKEDYFK	KIECFDSVEI	SGVEDRFNAS		
IKDKDFLDNE	ENEDILEDIV	LTLTLFEDRE	MIEERLKTYA		
LKRRRYTGWG	RLSRKLINGI	RDKQSGKTIL	DFLKSDGFAN	RNFMQLIHDD	700
SLTFKEDIQK	AQVSGQGDSL	HEHIANLAGS	PAIKKGILQT	VKVVDDELVKV	
MGRHKPENIV	IEARENQTT	QKGQKNSRER	MKRIEEGIKE	LGSQILKEHP	800
VENTQLQNEK	LYLYYLQNGR	DMYVDQELDI	NRLSDYDVDH	IVPQSFLKDD	
SIDNKVLTRS	DKNRGKSDNV	PSEEVVKMK	NYWRQLLNAK	LITQRKFDNL	900
TKAERGGLSE	LDKAGFIKRQ	LVETRQITKH	VAQILDSRMN	TKYDENDKLI	
REVKVITLKS	KLVSDFRKDF	QFYKVVREINN	YHHAHDAYLN	AVVGTALIKK	1000
YPKLESEFVY	GDYKVYDVRK	MIAKSEQEIG	KATAKYFFYS	NIMNFFKTEI	
TLANGEIRKR	PLIETNGETG	EIVWDKGRDF	ATVRKVLSMP	QVNIVKKTEV	1100
QTGGFSKESI	LPKRNSDKLI	ARKKDWDPPK	YGGFDSPTVA	YSVLVVAKVE	
KGKSKKLKSV	KELLGITIME	RSSFENPID	FLEAKGYKEV	KKDIIKLPK	1200
YSLFELENGR	KRMLASAGEL	QKGNELALPS	KYVNFLYLAS	HYEKLKGSPE	
DNEQKQLFVE	QHKHYLDEII	EQISEFSKRV	ILADANLDKV	LSAYNKHRDK	1300
PIREQAENII	HLFTLTNLGA	PAAFKYFDTT	IDRKRYTSTK	EVLDATLIHQ	
SITGLYETRI	DLSQLGGD				

Which Strategy will you adopt?

Blast?
GenePast?
Motif?

>15k results

A set of 6 mutations can convert SpCas9 into a nicking enzyme

(Known information used to illustrate a sequence search)

>WT_CRISPR-associated endonuclease Cas9

MDKKYSIGL	D	IGTN	SVGWAV	ITD	EYKVPSK	KFKVLGNTDR	50
LLFDSGET	T	ATRLK	RTARR	RYT	RRKN	RIC	100
LEESFLVEED		KKHERHPIFG		NIVDEVAYHE			200
LRLIYLALAH		MIKF	RGHFLI	EGDLNPDNSD			300
INAS	G	VDAKA	ILSARLSKSR	RLENLIAQLP	GEK	KNGLFGM	400
NFKSNFDLAE		DAKLQLSKDT		YDDDLNLLA			500
LLSDILR	V	N	ITKAPLSAS	MIKRYDEHHQ			600
FFDQSKNGYA		GYIDGGASQE		EFYKFIKPIL			700
KQRTFDNGSI		PHQIHLGELH		AILRRQEDFY			800
YVGPLARGNS		RFAWMTRKSE		ETITPWNFEE			900
NLPNEKVLPK		HSLLYEYFTV		YNELTKVKYV			1000
LLFKTNRKVT		VKQLKEDYFK		KIECFDSVEI			1100
IKDKDFLDNE		ENEDILEDIV		LTLTLFEDRE			1200
LKRRRYTGWG		RLSRKLINGI		RDKQSGKTIL			1300
SLTFKEDIQK		AQVSGQGDSL		HEHIANLAGS			
MGRHKPENIV		I	EMARENQTT	QKGQKNSRER			
VENTQLQNEK		LYLYYLQNGR		DMYVDQELDI			
SID	N	KVLTRS	DK	NRGKS	D	NV	
TKAERGGLSE		LDKAGFIKRQ		LVETRQITKH			
REVKVITLKS		KLVSDFRKDF		QFYKVVREINN			
YPKLESEFVY		GDYKVYDVRK		M	I	A	
TLANGEIRKR		PLIETNGETG		EIVWDKGRDF			
QTGGFSKESI		LPKRNSDKLI		ARKKDWDPPK			
KGKSKKLKSV		KELLG	I	TIME			
YSLFELENGR		KRMLASAGEL		QKGNELALPS			
DNEQKQLFVE		QHKHYLDEII		EQISEFSKRV			
PIREQAENII		HLFTLTNLGA		P	A	F	
SITGLYETRI		DLSQLGGD					

D10A

E762A

H840A

N854A

N863A

D986

Which Strategy will you adopt?

~~Blast?~~

Genepast?

Motif?

GenePast 90%

AlignmentPatentSubject AnnotationSubject SequenceQuery SequenceReport data issue

See all subjects mapped to this query

See all queries mapped to this subject

Your query contains this sequence. GQ subject-centric view.

Align len= 1157 aa, Errors= 51, Identity= 95.59%, Similarity= 97.32%

Query (WT_CRISPR-associated) len= 1368 aa, pos= 1-1156 aa, Identity query= 80.85%, Nb gaps query= 1, Alignment coverage 80.85%

Subject (US20160024568-1187) len= 1155 aa, pos= 1-1155 aa, Identity subject= 95.76%, Nb gaps subject= 2, Alignment coverage 95.76%

Q:1MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE60

S:1MDKKYSIGLDIGTNSVGWAVITDDYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE60

Q:61ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFG120

S:61ATRLKRSARRRYTRRKNRIRYLQEIFSSEMSKVDDSFHRLEESFLVEEDKKHERHPIFG120

Q:121NIVDEVAYHEKYPTIYHLRKKLVDSSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSD180

S:121NIVDEVAYHEKYPTIYHLRKKLVDSSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSD180

Q:181VDKLFQILVQTYNQLFEEPNASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGN240

S:181MDKLFQILVQTYNQLFEEPNASRVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGN240

Q:241LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI300

S:241LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI300

Q:301LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA360

S:301LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA360

Q:361GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH420

S:361GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH420

Q:421AILRRQEDFYFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE480

S:421AILRRQEDFYFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE480

Q:481VVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAFL540

S:481VVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPEFL540

1.1k results

A set of 6 mutations can convert SpCas9 into a nicking enzyme

(Known information used to illustrate a sequence search)

>WT_CRISPR-associated endonuclease Cas9

MDKKY	[^D]	GTNSVGWAV	ITDEYKVPSK	KFKVLGNTDR		
LLFDS		TRLKRTARR	RYTRRKNRIC	YLQEIFSNE	D10A	100
LEESFLVEED		KKHERHPIFG	NIVDEVAYHE	KYPTIYHLRK	E762A	200
LRLIYLALAH		MIKFRGHFLI	EGDLNPDNSD	VDKLFQILVC	H840A	300
INASGVDAKA		ILSARLSKSR	RLENLIAQLP	GEKKNGLFGM	N854A	400
NFKSNFDLAE		DAKLQLSKDT	YDDDLNLLA	QIGDQYADLF	N863A	500
LLSDILRVNT		EITKAPLSAS	MIKRYDEHHQ	DLTLLKALVR	D986	600
FFDQSKNGYA		GYIDGGASQE	EFYKFIKPIL	EKMDGTEELL		
KQRTFDNGSI		PHQIHLGELH	AILRRQEDFY	PFLKDNREKI		
YVGPLARGNS		RFAWMTRKSE	ETITPWNFEE	VVDKGASASQ		
NLPNEKVLPK		HSLLYEYFTV	YNELTKVKYV	TEGMRKPAFL		
LLFKTNRKVT		VKQLKEDYFK	KIECFDSVEI	SGVEDRFNAS		
IKDKDFLDNE		ENEDILEDIV	LTLTLFEDRE	MIEERLKTYA		
LKRRRYTGWG		RLSRKLINGI	RDKQSGKTIL	DFLKSDGFAN		700
SLTFKEDIQ		QGDSL	HEHIANLAGS	PAIKKGILQT		
MGRHKPENI	[^E]	ENQTT	QKGQKNSRER	MKRIEE		800
YENTQNEK		YNGR	DMYVDQELDI	NRLSDY	[^H]	
[^N]		TRSDNV	PSEEVVKKMK	NYWRQLENAR		900
[^N]		SEKRQ	LVETRQITKH	VACTLRPN		
REVKVITLKS		KLVSDFRKDF	QFYKVVREINN	YH	[^D]	1000
YPKLESEFVY		GDYKVYDVRK	MIAKSEQEIG	KA		
TLANGEIRKR		PLIETNGETG	EIVWDKGRDF	ATVRKVLSMP		1100
QTGGFSKESI		LPKRNSDKLI	ARKKDWDPPK	YGGFDSPTVA		
KGKSKKLKSV		KELLGITIME	RSSFENPID	FLEAKGYKEV		1200
YSLFELENGR		KRMLASAGEL	QKGNELALPS	KYVNFLYLAS		
DNEQKQLFVE		QHKHYLDEII	EQISEFSKRV	ILADANLDKV		1300
PIREQAENII		HLFTLTNLGA	PAAFKYFDTT	IDRKRYTSTK		
SITGLYETRI		DLSQLGGD		EVLDATLIHQ		

Which Strategy will you adopt?

~~Blast?~~

~~Genepast?~~

Motif?

A set of 6 mutations can convert SpCas9 into a nicking enzyme

(Known information used to illustrate a sequence search)

>WT_CRISPR-associated endonuclease Cas9

MDKKYSIGI	IGTNSVGWAV	ITDEYKVPSK	KFKVLGNTDR		
LLFDSGETAE	ATRLKRTARR	RYTRRKNRIC	YLQEIFSNE	D10A	100
LEESFLVEED	KKHERHPIFG	NIVDEVAYHE	KYPTIYHLRK	E762A	200
LRLIYLALAH	MIKFRGHFLI	EGDLNPDNSD	VDKLFQILVC	H840A	300
INASGVDAKA	ILSARLSKSR	RLENLIAQLP	GEKKNGLFGM	N854A	400
NFKSNFDLAE	DAKLQLSKDT	YDDDLNLLA	QIGDQYADLF	N863A	500
LLSDILRVNT	EITKAPLSAS	MIKRYDEHHQ	DLTLLKALVR	D986	600
FFDQSKNGYA	GYIDGGASQE	EFYKFIKPIL	EKMDGTEELL		
KQRTFDNGSI	PHQIHLGELH	AILRRQEDFY	PFLKDNREKI		
YVGPLARGNS	RFAWMTRKSE	ETITPWNFEE	VVDKGASASQ		
NLPNEKVLPK	HSLLYEYFTV	YNELTKVKYV	TEGMRKPAFL		
LLFKTNRKVT	VKQLKEDYFK	KIECFDSVEI	SGVEDRFNAS		
IKDKDFLDNE	ENEDILEDIV	LTLTLFEDRE	MIEERLKTYA		
LKRRRYTGWG	PLSRKLINGI	RDKQSGKTIL	DFLKSDGFAN		700
SLTFKEDIQK	VSQGQDSL	HEHIANLAGS	PAIKKGILOT		
MGRHKPENIV	MARENQTT	QKGQKNSRER	MKRIEEGII		800
VENLQNEK	MAKNGR	DMYVDQELDI	NRLSDYDVL		
SILVLTRS	[^N]	PSEEVVKKMK	NYWROLNAR		900
TKAERGGLSE	KRQ	LVETRQITKH	VAQILRMN		
REVKVITLKS	KLVSDFRKDF	QFYKVVREINN	YHHAHYN		1000
YPKLESEFVY	GDYKVYDVRK	MIAKSEQEIG	KATAKYFFYS		
TLANGEIRKR	PLIETNGETG	EIVWDKGRDF	ATVRKVLSMP		1100
QTGGFSKESI	LPKRNSDKLI	ARKKDWDPPK	YGGFDSPTVA		
KGKSKKLKSV	KELLGITIME	RSSFENPID	FLEAKGYKEV		1200
YSLFELENGR	KRMLASAGEL	QKGNELALPS	KYVNFLYLAS		
DNEQKQLFVE	QHKHYLDEII	EQISEFSKRV	ILADANLDKV		1300
PIREQAENII	HLFTLTNLGA	PAAFKYFDTT	IDRKRYTSTK		
SITGLYETRI	DLSQLGGD				

Which Strategy will you adopt?

~~Blast?~~

~~Genepast?~~

Motif?

A set of 6 mutations can convert SpCas9 into a nicking enzyme

(Known information used to illustrate a sequence search)

>WT_CRISPR-associated endonuclease Cas9

MDKKY	[^D]	GTN	SVGWAV	ITD	EYKVESK	KFKVLGNTDR	50
LLFDS		TRL	RTARR		RYTRRNKRIC	YLQEIFSNE	100
LEESFLVEED		KKHE	PHPIFG		NIVDEVAYHE	KYPTIYHLRK	
LRLIYLALAH		MIKF	RGHFLI		EGDLNPDNSD	VDKIFIQLVC	200
INAS	G	VDAKA	ILSARLSKSR		RLENLIAQLP	GEKKNGLFGM	
NFKSNFDIAE		DAKLQLSKDT	YDDDLNLLA		QIGDQYADLF		300
LLSDILRV		VNT	EITKAPLSAS		MIKRYDEHHQ	DLTLLKALVR	
FFDQSKNGYA		GYIDGGASQE	EFYKFIKPIL		EKMDGTEELL		400
KQRTFDNGSI		PHQIHLGELH	AILRRQEDFY		PFLKDNREKI		
YVGPLARGNS		RFAWMTRKSE	ETITPWNFEE		VVDKGASASQ		500
NLPNEKVLPK		HSLLYEYFTV	YNELTKVKYV		TEGMRKPAFL		
LLFKTNRKVT		VKQLKEDYFK	KIECFDSVEI		SGVEDRFNAS		600
IKDKDFLDNE		ENEDILEDIV	LTLTLFEDRE		MIEERLKTYA		
LKRRRYTGWG		RLSRKLINGI	RDKQSGKTIL		DFLKSDGFAN	RNFMQLIHDD	700
SLTFKEDIQ		QGDSL	HEHIANLAGS		PAIKKGILQT	VKVVDDELVKV	
MGRHKPENI		ENOTT	QKGQKNSRER		MKRIEE	SQILKEHP	800
VENTONER		ENGR	DMYVDQELDI		NRLSDY	PQSFLKDD	
[^N]		TRSR	[^N]		PSEEVVKKMK	NYWRQLENAR	900
SE		KRQ	LYETRQITKH		VACTLRPN	TKYDENDKLI	
REVKVITLKS		KLVSDFRKDF	QYKVVREINN		YHI	AVVGTALIKK	1000
YPKLESEFVY		GDYKVYDVRK	MAKSEQEIG		KAT	NIMNFFKTEI	
TLANGEIRKR		PLIETNGETG	EIVWDKGRDF		ATVRKVLSP	QVNIVKKTEV	1100
QTGGFSKESI		LPKRNSDKLI	ARKKDWDPKK		YGGFDSPTVA	SVLVVAKVE	
KGKSKKLKSV		KELLGITIME	RSSFENPID		FLEAKGYKEV	KKDLIIKLPK	1200
YSLFELENGR		KRMLASAGEL	QKGNELALPS		KYVNFLYLAS	HYEKLKGSPE	
DNEQKQLFVE		QHKHYLDEII	EIIEFSKRV		ILADANLDKV	LSAYNKURDK	1300
PIREQAENII		HLFTLTNLGA	PAAFKYFDTT		IDRKRYTSTK	EVLDATLIHQ	
SITGLYETRI		DLSQLGGD					

Which Strategy will you adopt?

~~Blast?~~

~~Genepast?~~

Motif?

A set of 6 mutations can convert SpCas9 into a nicking enzyme

(Known information used to illustrate a sequence search)

>WT_CRISPR-associated endonuclease Cas9

MDKKY [^D] GTN

D10A
E762A
H840A
N854A
N863A
D986A

Which Strategy will you adopt?

KPENIY [^E] ENQTT
[^N] TRS [^N] NV PSE

RLSDY [^H] PQSF
NN YHI [^D] N AVVGT

~~Blast?~~

~~Genepast?~~

Motif?

A set of 6 mutations can convert SpCas9 into a nicking enzyme

(Known information used to illustrate a sequence search)

>WT_CRISPR-associated endonuclease Cas9

MDKKYSIQ A IGTN

D10A
E762A
H840A
N854A
N863A
D986A

Which Strategy will you adopt?

KPENIV A MARENQTT
S A KVLTRS D A GKSDNV PSE

RLSDYDV A IVPQSF
NN YH A NYLN AVVGT

~~Blast?~~

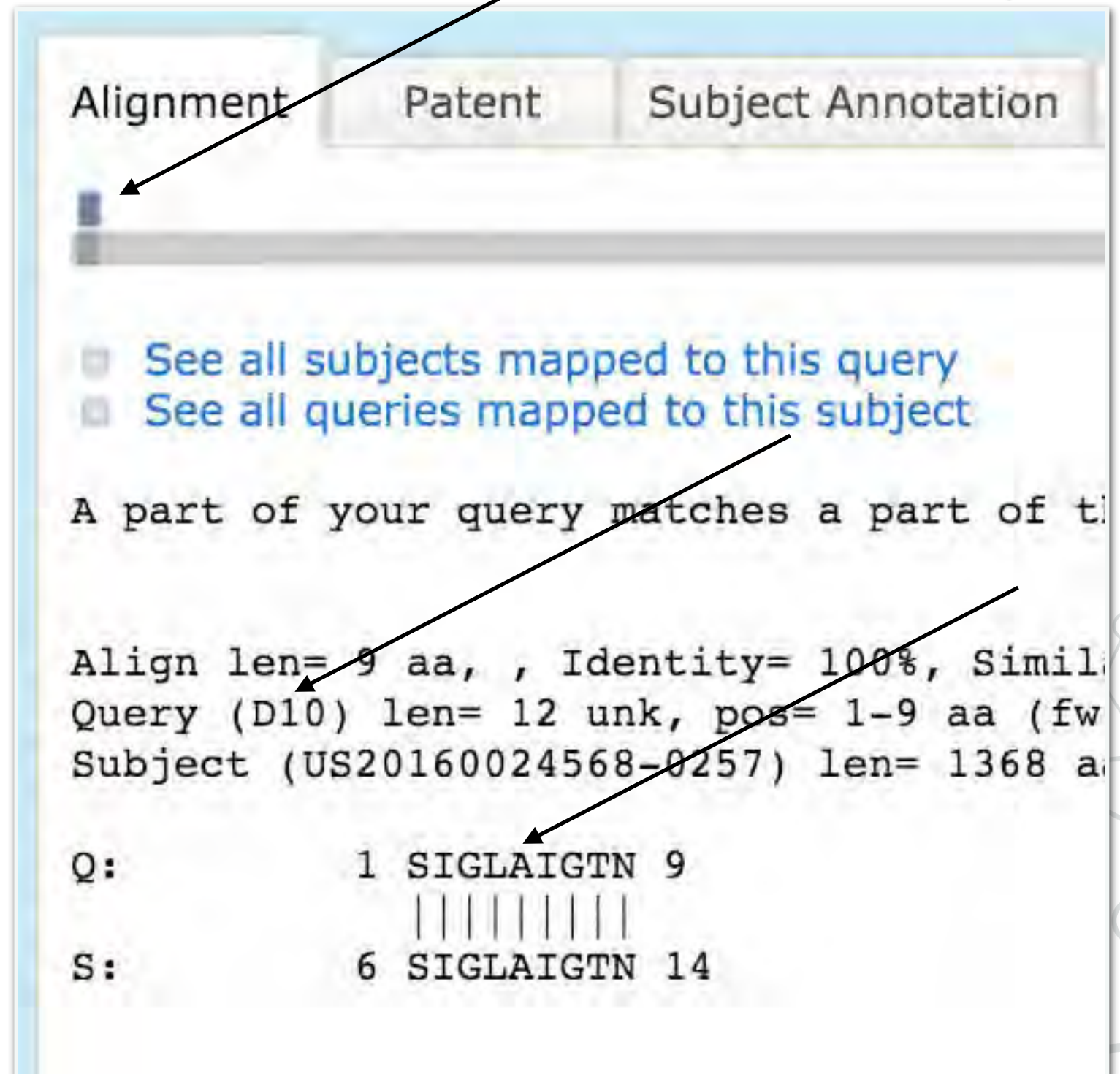
~~Genepast?~~

Motif?

6 Motifs

```
>D10
SIGL[^D]IGTN
>E762
NIVI[^E]MAREN
>H840
DVD[^H]IVPQ
>N854
DSID[^N]KVLT
>N863
RSDK[^N]RGKS
>D986
AH[^D]AYLNAV
```

< 1k results



Alignment Patent Subject Annotation

[See all subjects mapped to this query](#)

[See all queries mapped to this subject](#)

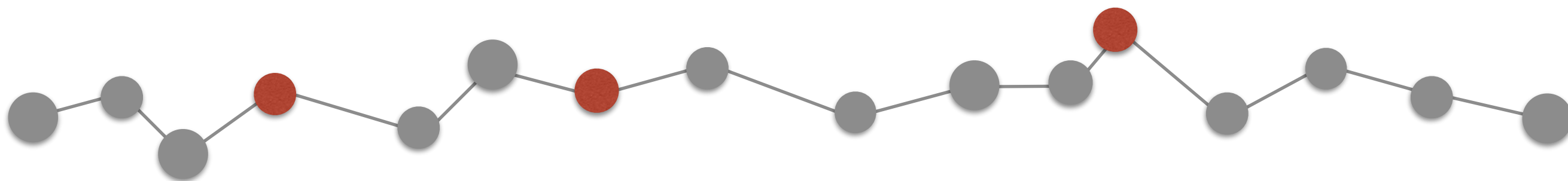
A part of your query matches a part of the

Align len= 9 aa, , Identity= 100%, Similarity= 100%, Query (D10) len= 12 unk, pos= 1-9 aa (fw), Subject (US20160024568-0257) len= 1368 aa

Q: 1 SIGLAIGTN 9

S: 6 SIGLAIGTN 14

Analysis



Find the PNs matching with multiple queries



Table of Contents

[Global Statistics](#)

[Number of documents per query](#)

[Number of documents by authority](#)

Global Statistics

There were original queries of which 6 have hits

There are 838 hits to the patent databanks.

There is a total of 144 patent documents from 119 INPADOC patent families.

The earliest priority date is : US20120216318 19990506

Number of documents per query

[Spreadsheet](#) of all hits classified by patents.

18 documents match 6 of the 6 queries

PATENT NUMBER	NB QUERIES	QUERY 1	QUERY 2	QUERY 3	QUERY 4	QUERY 5	QUERY 6	PATTERN
CA2894668	6	D10	D986	E762	H840	N854	N863	YYYYYY
EP2840140	6	D10	D986	E762	H840	N854	N863	YYYYYY
EP2848690	6	D10	D986	E762	H840	N854	N863	YYYYYY
KR20150105635	6	D10	D986	E762	H840	N854	N863	YYYYYY
US20140179006	6	D10	D986	E762	H840	N854	N863	YYYYYY
US20140189896	6	D10	D986	E762	H840	N854	N863	YYYYYY
US20140234972	6	D10	D986	E762	H840	N854	N863	YYYYYY
US20140248702	6	D10	D986	E762	H840	N854	N863	YYYYYY
US20140273231	6	D10	D986	E762	H840	N854	N863	YYYYYY
US20140310830	6	D10	D986	E762	H840	N854	N863	YYYYYY
US20150031134	6	D10	D986	E762	H840	N854	N863	YYYYYY
US20150079681	6	D10	D986	E762	H840	N854	N863	YYYYYY
US8795965	6	D10	D986	E762	H840	N854	N863	YYYYYY
US8871445	6	D10	D986	E762	H840	N854	N863	YYYYYY
US8889356	6	D10	D986	E762	H840	N854	N863	YYYYYY
US8932814	6	D10	D986	E762	H840	N854	N863	YYYYYY
WO2014093595	6	D10	D986	E762	H840	N854	N863	YYYYYY
WO2014093694	6	D10	D986	E762	H840	N854	N863	YYYYYY

25 documents match 5 of the 6 queries

PATENT NUMBER	NB QUERIES	QUERY 1	QUERY 2	QUERY 3	QUERY 4	QUERY 5	QUERY 6	PATTERN
KR20150105634	5	-	D986	E762	H840	N854	N863	YYYYYY
KR20150105956	5	-	D986	E762	H840	N854	N863	YYYYYY
US20140179770	5	-	D986	E762	H840	N854	N863	YYYYYY
US20140186919	5	-	D986	E762	H840	N854	N863	YYYYYY
US20140242699	5	-	D986	E762	H840	N854	N863	YYYYYY
US20140242700	5	-	D986	E762	H840	N854	N863	YYYYYY
US20140273234	5	-	D986	E762	H840	N854	N863	YYYYYY
US20140335620	5	-	D986	E762	H840	N854	N863	YYYYYY

See All Queries Mapped to a Subject

Result View Export Applications

Filtering Grouping Sorting Sharing Export PN List LifeQuest Full-text & Alignments

Grouped by Patent

Query % Id	Sub %
US201600245	100 0.6

Group by Subject and show 1 result per group.

Show only groups with

Group size greater than or equal to 3

Reset Apply

Export Applications

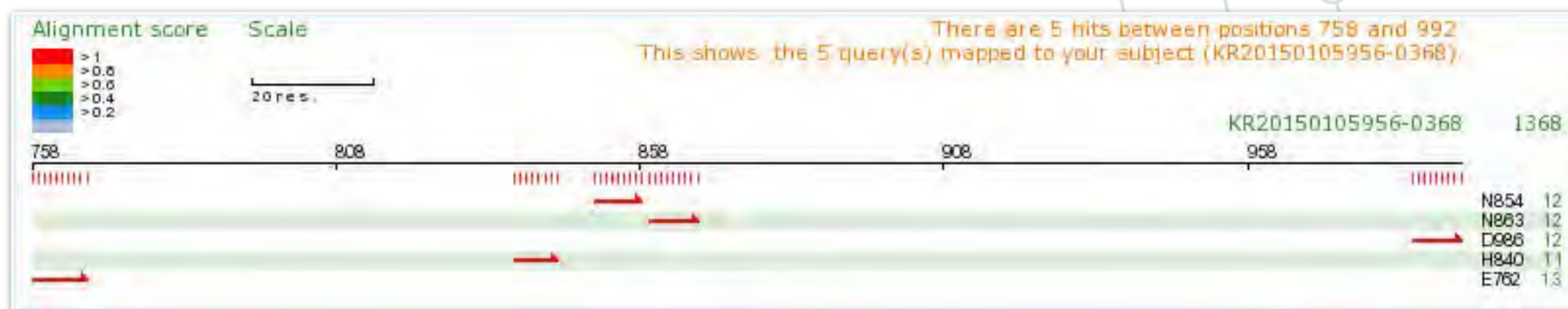
Grouping Sorting Sharing Export PN List LifeQuest Full-text & Alignments

Group by Subject and show All results per group.

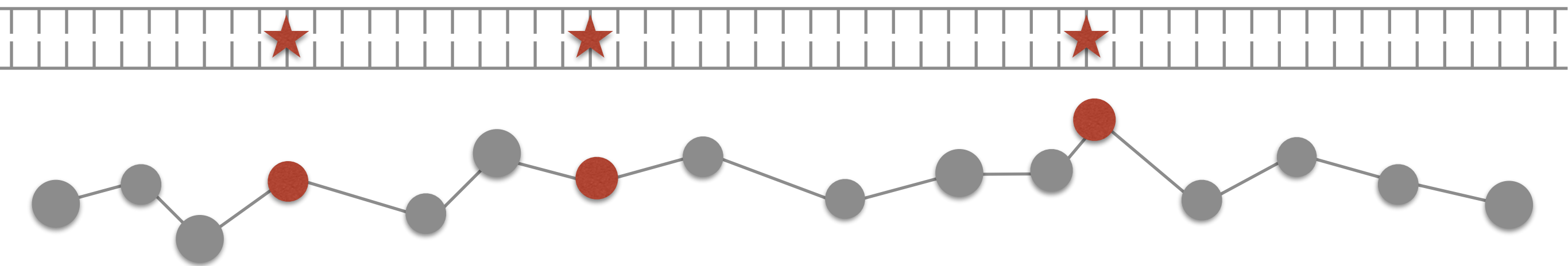
Show only groups with

Query Identifier	every member matches	D10	[+]	-	+
Query Identifier	every member matches	N864	[+]	-	+
Query Identifier	every member matches	E762	[+]	-	+

Reset Apply



Solutions (tricks?)



GenomeQuest

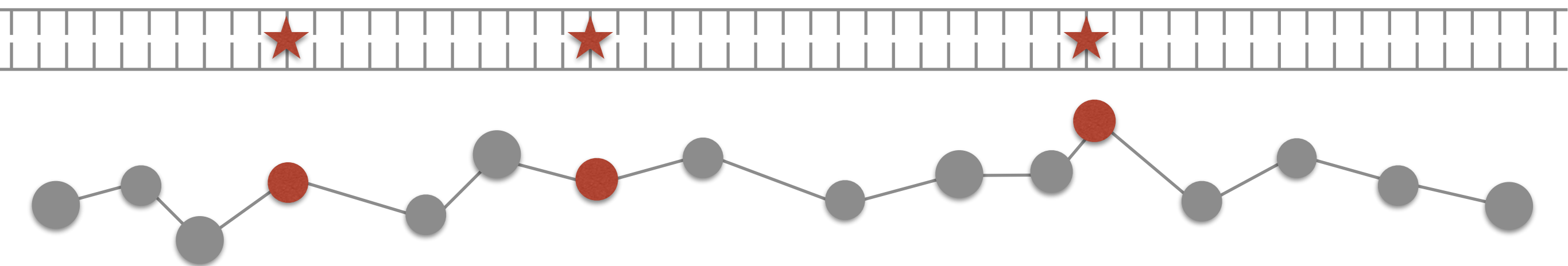
1. Find patents describing the Full length protein sequence - GenePast 90%
2. Find target mutations using short fragments overlapping the position of interest - Motif
3. Smart filtering / patent statistics

LifeQuest

1. Compare sequence searches results
2. Compare KW and Sequence search results

☒		WO2015157070A3	EDITAS MEDICINE	2015-Dec-30	CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING CYSTIC FIBROSIS	<table><tr><th>GQ Workflow</th><th>Query</th><th>Q%Id</th><th>S%Id</th><th>Aln%Id</th></tr><tr><td>WT cas9 90% genepast</td><td>WT_CRISPR-associated</td><td>100.0</td><td>100.0</td><td>100.0</td></tr><tr><td>WT cas9 90% genepast</td><td>WT_CRISPR-associated</td><td>100.0</td><td>100.0</td><td>100.0</td></tr><tr><td>WT cas9 90% genepast</td><td>WT_CRISPR-associated</td><td>100.0</td><td>100.0</td><td>100.0</td></tr><tr><td>WT cas9 90% genepast</td><td>WT_CRISPR-associated</td><td>100.0</td><td>100.0</td><td>100.0</td></tr><tr><td>WT cas9 90% genepast</td><td>WT_CRISPR-associated</td><td>99.93</td><td>100.0</td><td>100.0</td></tr><tr><td>WT cas9 90% genepast</td><td>WT_CRISPR-associated</td><td>1.97</td><td>100.0</td><td>100.0</td></tr><tr><td>WT cas9 90% genepast</td><td>WT_CRISPR-associated</td><td>1.97</td><td>100.0</td><td>100.0</td></tr></table>	GQ Workflow	Query	Q%Id	S%Id	Aln%Id	WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0	WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0	WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0	WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0	WT cas9 90% genepast	WT_CRISPR-associated	99.93	100.0	100.0	WT cas9 90% genepast	WT_CRISPR-associated	1.97	100.0	100.0	WT cas9 90% genepast	WT_CRISPR-associated	1.97	100.0	100.0
GQ Workflow	Query	Q%Id	S%Id	Aln%Id																																										
WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0																																										
WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0																																										
WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0																																										
WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0																																										
WT cas9 90% genepast	WT_CRISPR-associated	99.93	100.0	100.0																																										
WT cas9 90% genepast	WT_CRISPR-associated	1.97	100.0	100.0																																										
WT cas9 90% genepast	WT_CRISPR-associated	1.97	100.0	100.0																																										
☒		US20150368304A1	NOVARTIS	2015-Dec-24	NUCLEIC ACIDS AND PROTEINS FROM STREPTOCOCCUS GROUPS A AND B	<table><tr><th>GQ Workflow</th><th>Query</th><th>Q%Id</th><th>S%Id</th><th>Aln%Id</th></tr><tr><td>WT cas9 90% genepast</td><td>WT_CRISPR-associated</td><td>100.0</td><td>100.0</td><td>100.0</td></tr></table>	GQ Workflow	Query	Q%Id	S%Id	Aln%Id	WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0																														
GQ Workflow	Query	Q%Id	S%Id	Aln%Id																																										
WT cas9 90% genepast	WT_CRISPR-associated	100.0	100.0	100.0																																										
☒		WO2014204724A8	BROAD INSTITUTE HARVARD UNIVERSITY	2015-Dec-23	DELIVERY, ENGINEERING AND OPTIMIZATION OF TANDEM GUIDE SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION	<table><tr><th>GQ Workflow</th><th>Query</th><th>Q%Id</th><th>S%Id</th><th>Aln%Id</th></tr><tr><td>Postitions of interest - Motif</td><td>N854</td><td>100.0</td><td>0.66</td><td>100.0</td></tr><tr><td>Postitions of interest - Motif</td><td>N863</td><td>100.0</td><td>0.66</td><td>100.0</td></tr><tr><td>Postitions of interest - Motif</td><td>D986</td><td>100.0</td><td>0.66</td><td>100.0</td></tr><tr><td>Postitions of interest - Motif</td><td>H840</td><td>100.0</td><td>0.58</td><td>100.0</td></tr><tr><td>Postitions of interest - Motif</td><td>E762</td><td>100.0</td><td>0.73</td><td>100.0</td></tr><tr><td>WT cas9 90% genepast</td><td>WT_CRISPR-associated</td><td>99.63</td><td>99.63</td><td>99.63</td></tr></table>	GQ Workflow	Query	Q%Id	S%Id	Aln%Id	Postitions of interest - Motif	N854	100.0	0.66	100.0	Postitions of interest - Motif	N863	100.0	0.66	100.0	Postitions of interest - Motif	D986	100.0	0.66	100.0	Postitions of interest - Motif	H840	100.0	0.58	100.0	Postitions of interest - Motif	E762	100.0	0.73	100.0	WT cas9 90% genepast	WT_CRISPR-associated	99.63	99.63	99.63					
GQ Workflow	Query	Q%Id	S%Id	Aln%Id																																										
Postitions of interest - Motif	N854	100.0	0.66	100.0																																										
Postitions of interest - Motif	N863	100.0	0.66	100.0																																										
Postitions of interest - Motif	D986	100.0	0.66	100.0																																										
Postitions of interest - Motif	H840	100.0	0.58	100.0																																										
Postitions of interest - Motif	E762	100.0	0.73	100.0																																										
WT cas9 90% genepast	WT_CRISPR-associated	99.63	99.63	99.63																																										

Could this be made easier?



Hunting Single Point Mutations

PIUG BIOTECH 2017

Julien Doudement

Find sequences with a specific combination of variations

1

Enter
your Query sequence

Next

>WT_CRISPR-Associated Endonuclease Cas9

MDKKYSIGLD	IGTNSVGWAV	ITDEYKVPSK	KFKVLGNTDR	HSIKKNLIGA	
LLFDSGETAE	ATRLKRTARR	RYTRRKNRIC	YLQEIFSNE	AKVDDSFHR	100
LEESFLVEED	KKHERHPIFG	NIVDEVAYHE	KYPTIYHLRK	KLVDSTDKAD	
LRLIYLALAH	MIKFRGHFLI	EGDLNPDNSD	VDKLFQILVQ	TYNQLFEENP	200
INASGVDAKA	ILSARLSKSR	RLENLIAQLP	GEKKNGLFGN	LIALSLGLTP	
NFKSNFDLAE	DAKLQLSKDT	YDDDLNLLA	QIGDQYADLF	LAAKNLSDAI	300
LLSDILRVNT	EITKAPLSAS	MIKRYDEHHQ	DLTLLKALVR	QQLPEKYKEI	
FFDQSKNGYA	GYIDGGASQE	EFYKFIKPIL	EKMDGTEELL	VKLNREDLLR	400
KQRTFDNGSI	PHQIHLGELH	AILRRQEDFY	PFLKDNREKI	EKILTFRIPY	
YVGPLARGNS	RFAWMTRKSE	ETITPWNFEE	VVDKGASAQS	FIERMTNFDK	500
NLPNEKVLPK	HSLLYEYFTV	YNELTKVKYV	TEGMRKPAFL	SGEQKKAIVD	
LLFKTNRKVT	VKQLKEDYFK	KIECFDSVEI	SGVEDRFNAS	LGTYHDLLKI	600
IKDKDFLDNE	ENEDILEDIV	LTTLTFEDRE	MIEERLKTYA	HLFDDKVMKQ	
LKRRRYTGWG	RLSRKLINGI	RDKQSGKTIL	DFLKSDGFAN	RNFMQLIHDD	700
SLTFKEDIQK	AQVSGQGDSL	HEHIANLAGS	PAIKKGILQT	VKVVDLVKV	
MGRHKPENIV	IEMARENQTT	QKGQKNSRER	MKRIEEGIKE	LGSQILKEHP	800
VENTQLQNEK	LYLYYLQNGR	DMYVDQELDI	NRLSDYVDVH	IVPQSFLKDD	
SIDNKVLTRS	DKNRGKSDNV	PSEEVVKKMK	NYWRQLLNAK	LITQRKFDNL	900
TKAERGGLSE	LDKAGFIKRQ	LVETRQITKH	VAQILDSRMN	TKYDENDKLI	
REVKVITLKS	KLVSDFRKDF	QFYKVVREINN	YHHAHDAYLN	AVVGTAIIKK	1000
YPKLESEFVY	GDYKVYDVRK	MIAKSEQEIG	KATAKYFFYS	NIMNFFKTEI	
TLANGEIRKR	PLIETNGETG	EIVWDKGRDF	ATVRKVLSP	QVNIVKKTEV	1100
QTGGFSKESI	LPKRNSDKLI	ARKKDWDPPK	YGGFDSPTVA	YSVLVVAKVE	
KGKSKKLKSV	KELLGITIME	RSSFENPID	FLEAKGYKEV	KKDLIIKLPK	1200
YSLFELENGR	KRMLASAGEL	QKGNELALPS	KYVNFLYLAS	HYEKLKGSPE	
DNEQKQLFVE	QHKHYLDEII	EQISEFSKRV	ILADANLDKV	LSAYNKHARDK	1300
PIREQAENII	HLFTLTNLGA	PAAFKYFDTT	IDRKRYTSTK	EVLDTLIHQ	
SITGLYETRI	DLSQLGGD				

Find sequences with a specific combination of variations

2

Select the variations

Query	Subject is
D10	Anything Not D is one of A
E762	Anything Not E is one of A
H840	Anything Not H is one of A N
N854	Anything Not N is one of A T C P R F
N863	Anything Not N is one of A
D986	Anything Not D is one of A R

3

Browse the 24 sequences

>WT_CRISPR-Associated Endonuclease Cas9

```
MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA
LLFDSGETAE ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSSFFHR 100
LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLV DSTDKAD
LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP 200
INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP
NFKSNFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI 300
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI
FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR 400
KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY
YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFDK 500
NLPNEKVLPK HSLLEYEFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD
LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI 600
IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ
LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD 700
SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVDELVKV
MGRHKPENIV IEMARENQTT QKGQKNSRER MKRIEEGIKE LGSQILKEHP 800
VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVDH IVPQSFLKDD
SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL 900
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI
REVKVITLKS KLVSDFRKDF QFYKVVREINN YHHAHDAYLN AVVG TALIKK 1000
YPKLESEFVY GDYKVYDVRK MIAKSEQEIG KATAKYFFYS NIMNFFKTEI
TLANGEIRKR PLIETNGETG EIVWDKGRDF ATVRKVL SMP QVNIVKKTEV 1100
QTGGFSKESI LPKRNSDKLI ARKKDWDPPK YGGFDSPTVA YSVLVVAKVE
KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKD LIIKLPK 1200
YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE
DNEQKQLFVE QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKH RDK 1300
PIREQAENII HLFTLTNLGA PAAFKYFDTT IDRKRYTSTK EVLDATLIHQ
SITGLYETRI DLSQLGGD
```

PIUG BIOTECH 2017

Thank you!

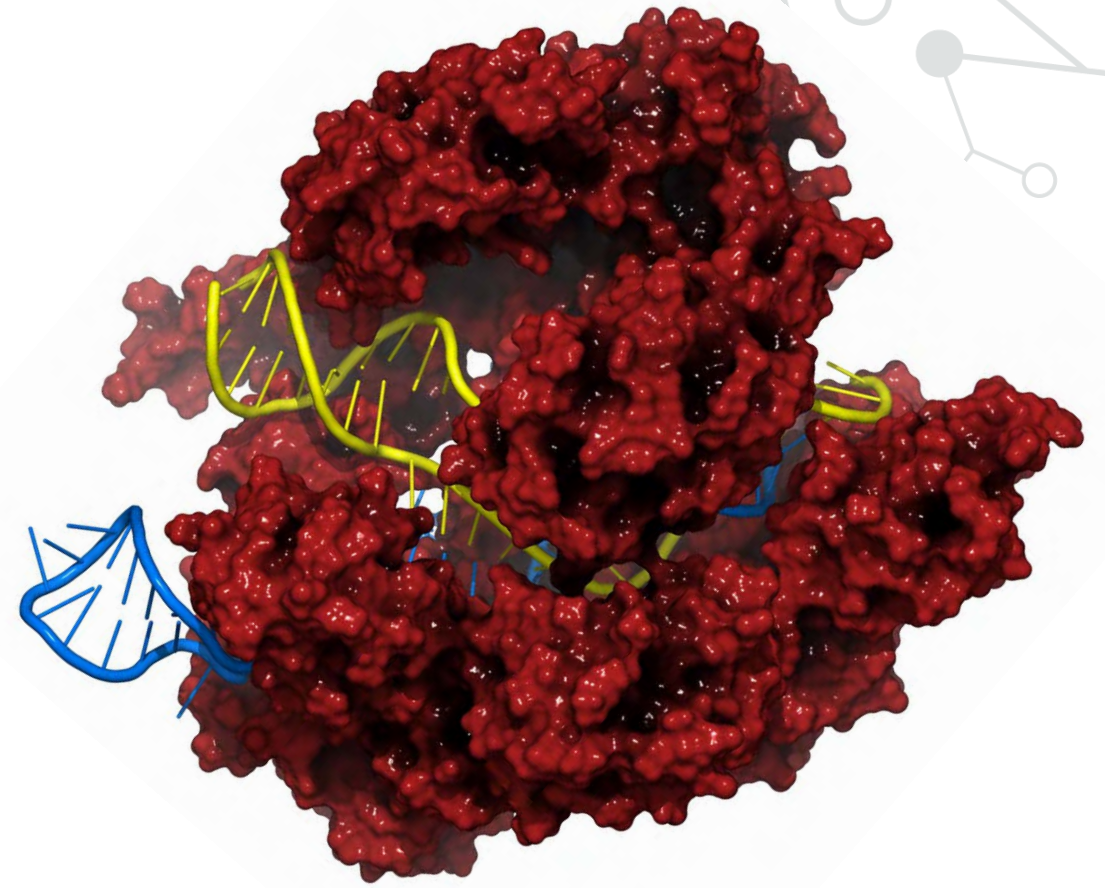
Julien Doudement

CRISPR / Cas9

New Challenges?



Cas9 Protein



Target DNA



crRNA



Replacement
DNA