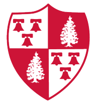
Montclair State University



Montclair State University Digital Commons

Theses, Dissertations and Culminating Projects

1-2024

Dynamic Genomic Evolution via Active LTR Transposable Elements in Maize

Taylor Isles

Follow this and additional works at: https://digitalcommons.montclair.edu/etd

Part of the Biology Commons

Abstract: Long terminal repeats (LTR) retrotransposons, found across eukaryotes, are transposable elements which can copy and insert themselves into other loci within a genome. These transposable elements are similar to retroviruses in that they rely on reverse transcriptase to "copy and paste" themselves elsewhere in the genome. From the RNA transcript they are able to use reverse transcriptase to make another DNA copy of themselves. This initially gave them the moniker, selfish genes. However, in the decades after the discovery of reverse transcriptase and LTR retrotransposons, it became known that non-genic DNA could have other functions. LTR retrotransposons are sources of mutation, inserting themselves into and mutating genes within their host organism. However, LTR retrotransposons are prone to mutation themselves, quickly becoming inactive, incapable of transposition. LTR retrotransposons constitute approximately 75% of the total genomic sequence of maize. The vast majority of these LTR retrotransposons are completely inactive. While these inactive retrotransposons can be useful for gleaning information about an organism's evolutionary history, the active LTR retrotransposons are of more interest. They are capable of causing mutations in current maize crosses and maize lines. In this study we created a tool that is adept at filtering out the inactive LTR retrotransposons and finding the location and sequence of potentially active LTR retrotransposons. Building off the freely accessible tool for locating LTR's, known as "LtrDetector", we added additional functionality. In an initial filter, LtrDetector found over three million potential LTR retrotransposons across thirty publicly available maize lines. After adding restrictions for size, an intact primary binding site, polypurine tract, reverse transcriptase, RNase and integrase we narrowed down that large pool of candidates to just twenty-seven potentially active LTR retrotransposons.

Dynamic Genomic Evolution via Active LTR Transposable Elements in Maize

By Taylor Isles

A Master's Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the degree of

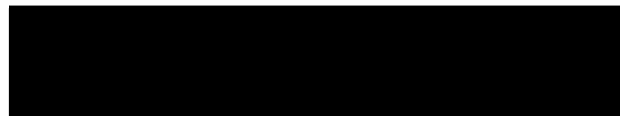
Master of Science

January 2024

College of Science and Mathematics

Department of Biology

Thesis Committee



Dr. Charles Du

Thesis Sponsor

Dr. John Gaynor

Dr. Carlos Molina

Committee Member

DYNAMIC GENOMIC EVOLUTION VIA ACTIVE LTR TRANSPOSABLE ELEMENTS IN MAIZE

A THESIS

Submitted in partial fulfillment of the requirements

For the Degree of Master of Science

By

Taylor Isles

Montclair State University

Montclair, NJ

2023

Copyright@2023 by Taylor Isles. All Rights Reserved.

ACKNOWLEDGEMENTS

I would like to extend a major thank you to my committee members Dr. Charles Du, Dr. John Gaynor and Dr. Carlos Molina for their assistance over the past year. I'd like to thank the Wehner committee for selecting me for the Wehner Student Research Program. I would also like to thank Angadh Singh for his help. Finally, I would like to thank Shay Wiggington, Istiaque Chowdhury, Dominick Altamura and Laura Lepore for their aid in the next stage of the project.

TABLE OF CONTENTS

Introduction……………………………………………………………………………… 9

Methods…………………………………………………………………………………..12

Results………………………………………………………………………………... 16

Discussion…………………………………………………………………………… 23

References…………………………………………………………………………… 26

Appendix A: Annotated Sequences…………………………………………………….. 29

Appendix B: PBS Logos…………………………………………………………... 70

Appendix C: Code…………………………………………..…………………... 72

LIST OF TABLES

Table 1: 1,792 Potentially Active LTR Retrotransposons after First Filter……….. 17

Table 2: LTR Retrotransposons Filtering for Conserved Domains………………... 18

Table 3: LTR Retrotransposons and Conserved Domains Found in Sequence…… 21

List of Figures

Figure 1: Domain Logos...…………………………...……………………….................. 19

Figure 2: CML 103 Chr.3 Annotation……...……………………………………….... 22

INTRODUCTION:

LTR retrotransposons are transposable elements that can be found throughout the genomes of eukaryotes. In some maize lines they comprise up to 75% of the entire genome (Baucom, et al, 2009). In contrast, the two other classes of transposable retroelements made up just 1% of the maize genome (Baucom et al., 2009). In other eukaryotes it makes up a much lower but still significant portion of the genome, about 8% in humans (Cordaux and Batzer, 2009). It was initially thought that the genome was static. Genes and the large regions of "junk" between them were believed to be in fixed positions (Goodier and Kazazian, 2008). Today it is known that this is not the case. There are many mobile elements in the genomes of eukaryotes, and many have not yet been properly identified. The work of identification started in the 1940's when Barbara McClintock discovered what she called "Controlling elements". She found two distinct elements in maize, Ds (Disassociation)elements and Ac(Activator) elements (Goodier and Kazazian, 2008). Thus, beginning a long relationship between maize research and mobile element research. These transposable elements began to be sequenced in the eighties, and they were characterized by a retroviral-like sequence. Mount and Rubin (1985) discovered homology between transposons and the retroviral protein domains for gag protein, protease, reverse transcriptase, and integrase. Where they differed from retroviruses was in the lack of an envelope protein (Eickbush, Jamburuthugoda, 2008). Further experiments demonstrated that these elements replicated themselves through reverse transcription (Boeke et al., 1985). They were dubbed retrotransposons. These transposable elements or transposons are divided into two large groups, DNA transposons and retrotransposons. DNA transposons use a "cut and paste" mechanism where DNA is

removed from one location in the genome and relocated to another area (Goodier, Kazazian, 2008). Retrotransposons use a "copy and paste" system. The retrotransposons transcribes an RNA copy before that copy is reverse transcribed back into the genome as DNA (Finnegan, 2012). In this way retrotransposons can cause the genome to expand (Finnegan, 2012). Retrotransposons are found in bacteria, archaea and eukaryotes, implying an ancient origin (Beauregard et al., 2008). Retrotransposons are then grouped into two large families, extrachromosomally-primed (EP) retrotransposons and target-primed (TP) retrotransposons (Beauregard et al., 2008). EP retrotransposons are transcribed into an RNA template and then reverse transcriptase synthesizes cDNA from the RNA template (Beauregard et al., 2008). Integrase then allows access to the target site and the dsDNA is spliced back into the genome (Beauregard et al., 2008). The majority of these EP retrotransposons are LTR retrotransposons (Beauregard et al., 2008). LTR stands for long terminal repeats. They are defined by these two identical sequences at the 5' and 3' ends of the transposon. (Mount and Rubin, 1985) Within this group there are multiple different LTR retrotransposons, the most researched being Type I (Copia) and Type III (gypsy). These two types are differentiated by the order of their domains. Copia LTR's have integrase as the third domain in the retrotransposon, while Gypsy LTR's have the integrase domain last (Eikbush and Jamburuthugoda, 2008). The other family of retrotransposons are known as target-primed retrotransposons. This family is largely comprised of non-LTR retrotransposons but there are several exceptions (Beauregard et al., 2008). TP retrotransposons are different from EP retrotransposons in that the RNA transcript binds to a 3' hydroxyl group at the target site after endonuclease nicks a strand (Beauregard et al., 2008). To simplify, EP retrotransposons synthesize cDNA and then

integrates into the genome, while TP retrotransposons integrate first and then synthesize cDNA from the attached RNA transcript (Beauregard et al., 2008). Regardless of the mechanisms they use to transpose themselves these transposable elements have been shown as the cause of mutation in multiple different eukaryotes (Dooner et al., 2019). The genes and regulatory elements affected by these mobile elements vary based on where they insert themselves, as they can insert themselves in or near genes (Dooner et al., 2019). This is part of why retrotransposons are an important driver of evolutionary changes. For example, LTR retrotransposons and other transposable elements were shown to grow exponentially in the hominid genome shortly before major speciation events (Kim et al., 2004). However, not all LTR retrotransposons are capable of producing mutations. Most LTR retrotransposons are inactive. Older LTR retrotransposons are less capable of transposition than their younger counterparts (Jedlicka et al., 2020). LTR retrotransposition is not a perfect process, around one-third of all transpositions result in a “broken”, non-standard LTR retrotransposon, which goes a long way in explaining why older LTR retrotransposons become incapable of transposition (Dooner et al., 2019). In order to better understand transposons, they first must be found, sequenced, and annotated in order to better understand their effects on the genome. The innovations of the past two decades in bioinformatics and sequencing technology have been immensely helpful in this aspect (Dooner et al., 2019). It has become much simpler to sequence entire genomes rather than focusing on single genes. In this project, we developed a tool to locate LTR retrotransposons, making it easier to identify potentially active, young retrotransposons from among the many inactive ones. To do this we searched for the hallmarks of LTR retrotransposons, two identical LTR’s, a

primary binding site, a polypurine tract and the conserved domains for protease, reverse transcriptase, RNase and integrase (Goodier and Kazazian 2008). This bioinformatic tool read through 30 different publicly available maize lines from Maize GDB (Woodhouse, 2021), and filtered out any that did not match the typical LTR retrotransposon sequence. We hope that this tool greatly eases the process of identifying active LTR retrotransposons in not only the maize genome, but many different plant genomes. The tool is largely species agnostic, however, if applied to a very different organism, like a mammal or bacteria, the exact protein sequences searched for would need to be changed. In addition, we also discuss the ongoing work to grow maize crosses and search for novel active LTR retrotransposons.

METHODS:

To identify all the potential active LTR retrotransposons we employed the use of the program LtrDetector. This program looks for sets of identical long terminal repeats (Valencia and Girgis, 2019). Once this program was integrated, we focused on refining the search, as at this point we had over 100,000 candidate LTR's. Next, we had to add parameters in order to search for other components that are part of active. First, we added a filter looking for Polypurine Tracts, a stretch of multiple purines that is adjacent to the 3' LTR. Then a filter was added to look for the primary binding site. We used multiple different PBS logos to capture as many permutations of the PBS as possible. Then we generated amino acid logos for reverse transcriptase, RNase and GAG proteins. The program converted each LTR retrotransposon to its corresponding amino acid sequence and compared it to the generated logos to confirm that the desired domains were present.

The program confirmed the right and left LTR were identical. It checked for the presence of identical TSDs. It also checked for a match to the primary binding site logo. Finally checking for a polypurine tract upstream of the 3'LTR. To accomplish these tasks, the tool first loaded the indexed position of all retrotransposons in the sample set provided by the LTR detector. However, since the detector only provided the positions of the LTR in the genome and not the sequences themselves, the next step must be to pull and read the sequences present at these locations. To ensure each retrotransposon was unique, they were labeled according to the chromosome they are located on, as well as the starting and ending base pair index. For example, a retrotransposon located on the 3rd chromosome from the 10th base pair to the 25th would be labeled as "chr3:10-25". This labeling method allowed for the use of a HashMap, which sped up the process. The tool then proceeded to cut out the 5' and 3' LTR from each sequence and check whether they are identical. Subsequently, the tool examines the 20 base pairs upstream of the 3' LTR to determine the number of purines present. A purine content of more than 85% in the 20 base pair sample would indicate the presence of a polypurine tract. The tool also checked for the presence of a primer binding site immediately following the 5' LTR while ensuring that the target site duplications are present and identical. Only sequences that meet all the previous criteria were eligible for the next filter. Next, the tool converted each sequence into its six possible reading frame amino acid chains. Each chain is then checked for the presence of Reverse Transcriptase, RNase, and GAG proteins using a pattern identifier known as regex. Finally, the list of candidates was written to an output.fa file, which can be manually annotated and confirmed as active LTR retrotransposons.

Annotation

After the data collection tool went through the files on the high-performance computing cluster (HPC). The HPC runs on Linux and is comprised of 55 computing nodes, 24 GPU's, 1416 CPU cores, and 12.5TB of aggregate RAM. We manually annotated the transposons in order to verify that the tool was working. In order to annotate the LTR retrotransposons we performed the following steps. The Long Terminal Repeats are two identical sequences at the 5' and 3' end of LTR retrotransposons. They should be right next to the TSD's (target site duplications) and they should be not be more than 1000 base pairs. In this way we used what we knew about one element to find the other. LTR's are always directly next to TSD's, so I knew that the bases right next to it are part of the LTR. I took the first 10 or so bases that are right next to the 5' TSD and searched the document for copies (control+f or command+f), and I found the beginning of the 3' LTR this way. The 3' LTR ran from this sequence to the 3' TSD, which gives us the entire LTR, then by knowing the sequence of the 3' LTR, which means I could find the sequence of your 5' LTR and can highlight it as well. After that I found the Polypurine Tract. This is a site of purines (G's and A's) that should be no more than 2 base pairs from the 3' LTR. It should be 5-8 base pairs long. It could be a little longer but it's usually between 5-8 base pairs. Finally, I located the Primary Binding Site. The primary binding site is located right next to the 5' LTR, usually not more than a few base pairs away from the 5'LTR. There is only one primary binding site per transposon and we used logos from previous studies in order to search and identify the correct sequence.

In the logo below, the larger the number, the more likely that nucleotidecan be found there. In the below example, a primary binding site will always start with TGGTA and

will usually be followed by T, but can be followed by C. After identifying these regions that are within each LTR we needed to identify the domains inside each retrotransposon. I initially looked for gag, protease, RNAase H and integrase domains. To find these we used NCBI (https://www.ncbi.nlm.nih.gov/) and Uniprot's (https://www.uniprot.org/) conserved domain database. Within these databases we found the preexisting logos for the domains we were looking for. We also generated our own logos by feeding sequences identified by the conserved domain viewer (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) on NCBI to the web app weblogo (https://weblogo.berkeley.edu/). This allowed us to confirm that the logos we selected were a match for the ones being used by the database to identify the domains in our sequences. Unfortunately, the gag logos that we generated and the logos used by the conserved domain database were very weak and were ultimately cut from our filter. The domain database noted that the gag domain logo was weak and not as strong as the other logos we were using. In order to find those other domains I entered each sequence from the first base pair after the PBS to the last base pair just before the PPT into the conserved domain viewer on NCBI (for this I kept the default settings). After the application was done processing the request it displayed our annotated sequence. This included domains mapped onto our sequence as well as their active sites, down to the nucleotide level. This permitted me find domains within our sequences manually. The viewer also provided links and logos for tagged domains allowing us to then incorporate NCBI's logos as well as start to develop our own using our annotated sequences.

RESULTS:

Bioinformatic Tool Filters Out Broken LTR's

LtrDetector initially detected 3,771,936 identical LTR's across 30 maize inbred lines. After filtering out the detected sequences that lacked a PBS, a PPT, two identical LTR's and were longer than 5,000 base pairs in length there were 1,792 detected LTR retrotransposons. The filter was not configured to search for conserved domains at this step. Table 1 displays the total number of detected LTR retrotransposons, the maize line in which they occurred and chromosome they can be found on. Most of the samples averaged between 30 to 45 detections. EP1, F7 and W22 are outliers with 297, 236 and 184 detections respectively. This consistency across multiple chromosomes is encouraging as a large difference between every maize line might imply that the quality of the genome sequencing was different between maize lines. A consistent number of LTR retrotransposon detections implies that the maize lines have undergone quality long read sequencing. PE0075 has only eight detected LTR retrotransposons. This is the lowest number of detections across the dataset and could be due to PE0075 having a lower quality genome sequencing. It also might imply PE0075 has a very low number of LTR retrotransposons. The LTR retrotransposons sampled from the 1,792 had all the identifiers we were looking for at this stage, indicating that our bioinformatic tool was performing correctly.

Table 1: 1,792 Potentially Active LTR Retrotransposons after First Filter

	chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	**Total**
B73	6	6	2	5	3	3	4	1	3	5	38
B97	**5**	**6**	**3**	**2**	**8**	**5**	**3**	**4**	**2**	**2**	40
CML103	4	5	5	4	7	2	5	4	2	2	40
CML228	5	4	4	2	10	3	3	3	2	3	39
CML247	6	3	5	3	7	6	3	5	0	3	41
CML277	3	4	4	3	9	3	4	6	3	3	42
CML322	5	4	5	4	7	5	1	1	3	3	38
CML333	6	3	4	2	8	3	1	5	2	3	37
CML69	1	6	4	2	6	2	4	3	1	3	32
DK105	5	4	4	7	6	10	4	6	7	3	56
EP1	30	40	45	35	29	27	25	24	21	21	297
F7	27	22	31	27	23	24	22	21	19	20	236
HP301	1	10	2	3	3	3	4	4	1	5	36
Ia453	6	3	0	2	7	4	5	5	6	1	39
IL14H	5	5	1	5	5	2	6	5	7	3	44
K0326Y	4	6	2	7	4	4	7	6	2	3	45
Ki11	8	3	3	2	6	5	4	8	2	3	44
Ki3	7	3	4	2	8	4	3	5	1	4	41
Ky21	5	7	1	7	3	5	2	2	5	4	41
M162W	3	7	3	0	6	3	4	3	3	8	40
M37W	8	6	3	7	3	4	2	7	2	1	43
Mo18W	5	4	4	2	8	3	3	4	1	5	39
Ms71	10	6	8	0	4	3	3	4	3	5	46
NC350	3	4	6	0	5	5	2	10	3	0	38
NC358	7	4	4	3	5	7	9	4	3	5	51
Oh43	8	3	5	2	4	3	5	3	3	5	41
PE0075	0	1	1	2	0	1	1	0	2	0	8
Tx303	3	7	3	2	5	4	4	4	1	3	36
Tzi8	4	4	4	5	5	7	3	1	5	2	40
W22	20	25	22	19	20	13	15	25	19	6	184
											1792

Table 1 - LTR retrotransposons detected on 30 maize lines from MaizeGDB (https://www.maizegdb.org/) and the chromosome each is located on. At this stage the filter only selected sequences that were less than 5,000 base pairs with identical LTR's, a polypurine tract and a Primary Binding Site. The total number of retrotransposons that met these criteria across all Maize lines was 1792 detected LTR retrotransposons, narrowed from 3,771,936. The three maize lines with the greatest number of detected retrotransposons were EP1 (297 retrotransposons), F7 (236 retrotransposons) and W22(184 retrotransposons). PE0075 had the lowest with eight, making the range of detections eight to 297 with a mean of 59.73.

Table 2: LTR Retrotransposons Filtering for Conserved Domains

Maize Cross	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10
B73				1	1					
B97									1	
CML103	1				1				1	
CML322			1		1					
CML333	1									
EP1				2			1			1
F7										1
HP301					1					
LA453					1				1	
IL14H					1				1	
K0326Y				1					1	
NC358	1				1					
Oh43					1					
PE0075				1						
Tx303									1	
Tzi8					1					

Table 2 – Active LTR retrotransposons detected by the bioinformatic tool after screening for Reverse Transcriptase, RNAase and Gag domains. The total number of detected LTR retrotransposons is twenty-seven. Maize lines that produced no hits were not included in the table.

Bioinformatic Tool Filters out Retrotransposons Lacking Desired Domains

After implementing filters for conserved sequences of domains only 27 LTR retrotransposons contained domains for reverse transcriptase, gag protein, and RNase. In order to filter the retrotransposons with the desired domains we generated logos from existing LTR retrotransposons on the NCBI database (Sayers et al., 2022). After finding LTR retrotransposons we used NCBI's domain database to identify the domains within these sequences, aligned them and used them to generate logos with the weblogo browser application (Crooks et al., 2004). The bioinformatic tool was then given pieces of these logos and filtered any sequence that did not contain them. In order to make the tool more expansive, multiple different sequences were added for reverse transcriptase and RNAase. As long as the tool found one of the sequences for reverse transcriptase it would

pass the filter. In this way we made sure we found as many permutations of active domains as possible. On Table 2, the 27 retrotransposons are listed with number, maize line and location within the genome. The filter did not find LTR retrotransposons in any of the other fourteen maize lines. After running the tool with filters for reverse transcriptase, RNase, and integrase it was decided that we would proceed without filtering for protease, as twenty-seven was a workable amount of retrotransposons for manual annotation.

Figure 1: Domain Logos

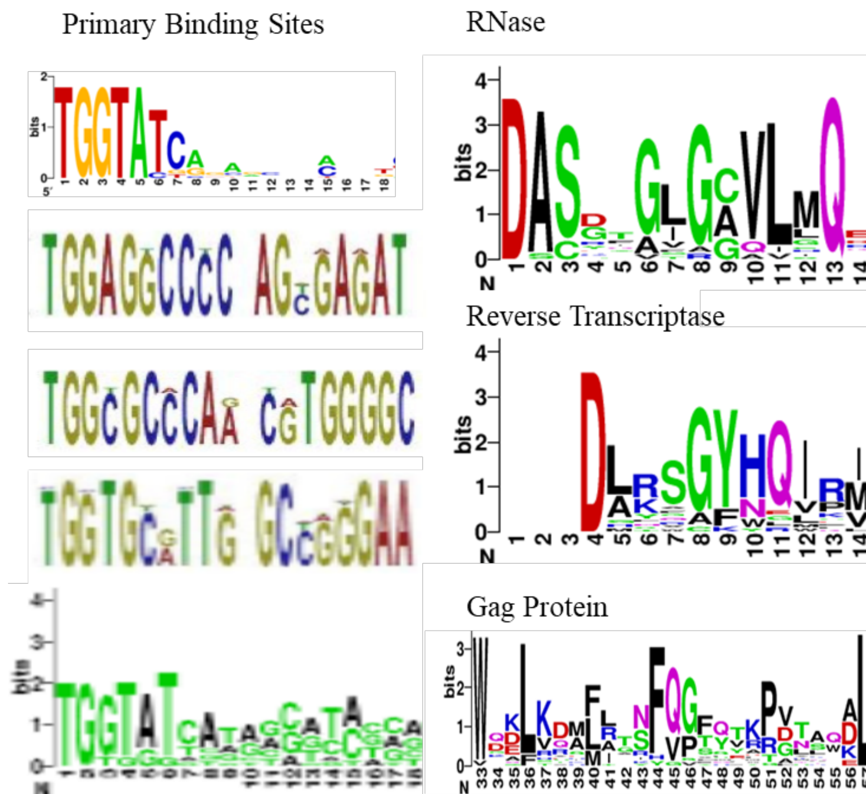


Figure 1 –The logos generated on weblogo.berkely.edu for the Primary Binding site of LTR retrotransposons, as well as the logos of RNase, reverse transcriptase and gag protein. The Gag protein had more variance than both RNase and Reverse Transcriptase, and was not annotated as it was difficult for our bioninformatic tool and NCBI's database to identify accurately in a sequence.

Table 3: LTR Retrotransposons and the Conserved Domains Found in Sequence

LTR Retrotransposons Conserved Domains				
LTR Retrotransposon	Protease	Reverse Transcriptase	Rnase	Integrase
B73 Ch4		X	X	X
B73 Chr 5		X	X	X
B97 Chr 9		X	X	X
CML103 Chr 1		X	X	X
CML103 Chr 5		X	X	X
CML103 Chr 9		X	X	
CML322 Chr 3	X	X	X	X
CML322 Chr 5		X	X	X
CML333 Chr 1				
EP1 Chr 4: 8015632-8020208	X	X	X	X
EP1 Chr 4: 258921835-258925715		X	X	X
EP1 Chr 7		X	X	X
EP1 Chr10		X	X	X
F7 Chr10		X	X	X
HPL301 Chr 5		X	X	X
LA453 Chr 5		X	X	X
LA453 Chr 9	X	X	X	X
IL14H Chr 5		X	X	X
IL14H Chr 9		X	X	X
K0326Y Chr 4		X	X	X
K0326Y Chr 9		X	X	
NC358 Chr 1	X	X	X	X
NC358 Chr 5		X	X	X
Oh43 Chr 5		X	X	X
PE0075 Chr 4		X	X	
Tx303 Chr 9	X	X	X	X
Tzi8 Chr 5		X	X	X

Table 3 - The Twenty-Seven LTR retrotransposons and their conserved domains. While not filtered for by the bioinformatic tool. Integrase was found in nearly all of sequences after the sequences were manually annotated.

Manual Annotation Confirms Presence of Complete LTR Retrotransposons

Within the annotated sequences almost all contained the requisite conserved domains. There were 27 total detected LTR retrotransposons. Twenty-six of them had Reverse Transcriptase and RNase and twenty-four of them had integrase. This was confirmed by checking against the logos of the conserved domain and pfam databases on NCBI and Uniprot. This shows that the bioinformatic tool was functioning correctly as most of LTR retrotransposons it found had the requested conserved domains. When selecting logos, we aimed to use active sites and the most highly conserved regions according to NCBI. We checked for protease manually in these twenty-seven sequences. Out of twenty-seven retrotransposons, six of these retrotransposons had a protease domain. They were located on chromosome 3 on maize line CML322, on chromosome 9 in maize line La453, on chromosome 1 in maize line NC358, on chromosome 9 in maize line Tx303, on chromosome 4 in maize line EP1 and on chromosome 4 in maize line B73. These six LTR retrotransposons show the least signs of degradation and mutation, making them the most likely to be active and capable of transposition. Given the order of the domains contained within the annotated sequences, (Figure 2) it is a type 3 LTR retrotransposon. Of note is that the incomplete LTR retrotransposons often repeated across maize lines. The sequences of the LTR retrotransposons found on chromosome 5 were extremely similar to one another. They had the same target site duplication CTCCA, the same three conserved domains, reverse transcriptase, RNase and integrase. The sequences on chromosome four and 9 often share repeating patterns as well.

Figure 2: CML322: Ch3 Annotated

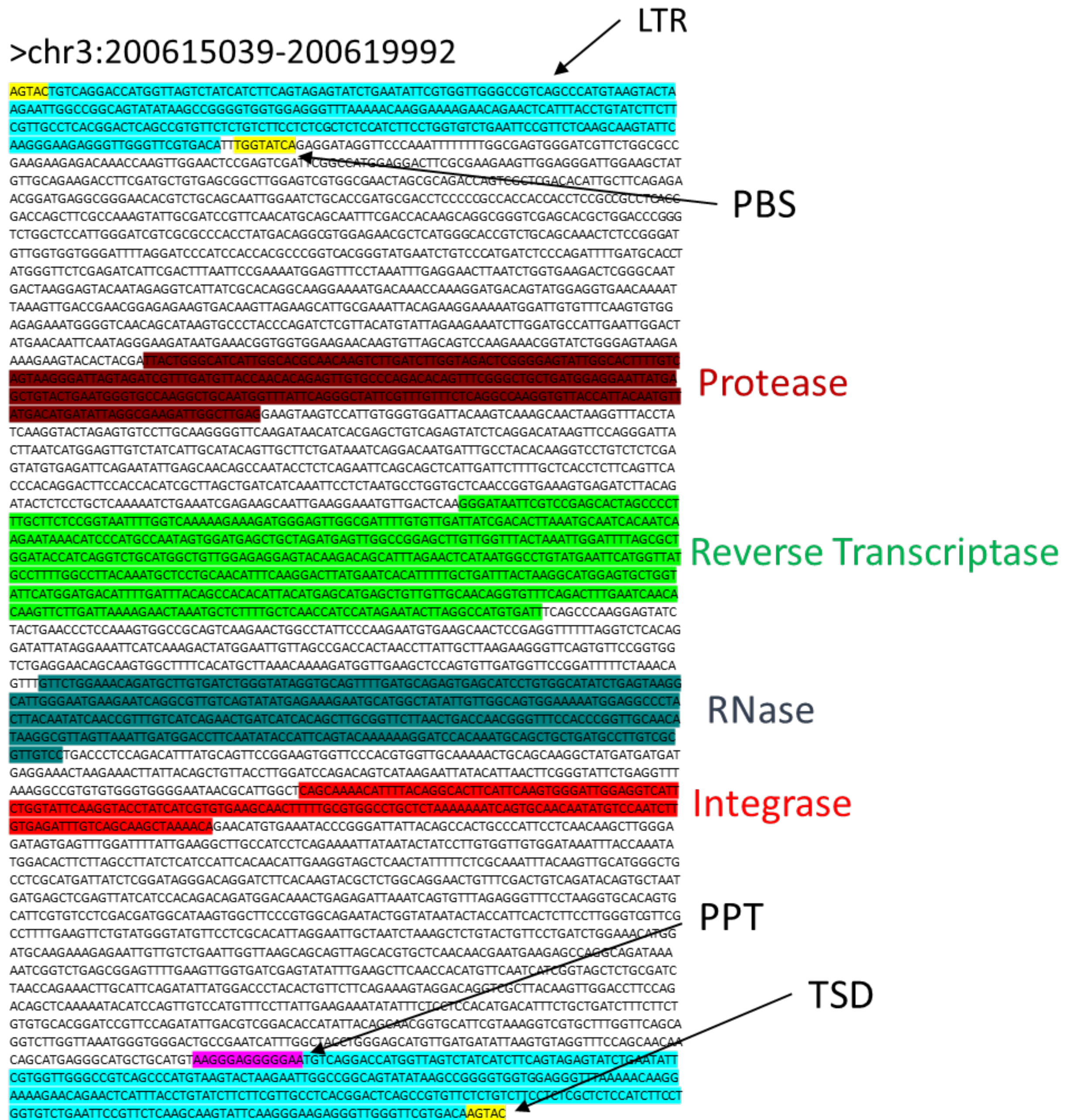


Figure 2 - An example of a fully annotated sequence pulled from the genome by the bioinformatics tool. This sequence has all the conserved domains requested. From top to bottom, in yellow – The First TSD, followed by the 5' LTR in sky blue. The primary binding site in yellow. Then after this the first filtered domain is reverse transcriptase, then RNase and integrase. Then is the PPT in purple. Finally the retrotransposon is We did not annotate gag due to the logo for gag being weakly conserved.

DISCUSSION:

LTR retrotransposons are retroviral-like transposable elements that are capable of driving mutation within eukaryotic genomes, especially plant genomes (Jedlicka et al., 2020). The active retrotransposons can be very low copy number and surrounded by inactive LTR retrotransposons, in repetitive regions, making finding them difficult (Dooner et al., 2019). Active LTR retrotransposons often adhere to a typical set of domains and features, two LTR's, a PBS, PPT and the conserved domains for gag, protease, reverse transcriptase, RNase and integrase (Goodier and Kazazian, 2008). The bioinformatic tool developed here will be useful in future projects due to its ability to detect domains analogous to those typical, functioning LTR retrotransposons. The tool was able to detect retrotransposons within the thirty publicly available maize lines. These maize lines were sequenced accurately enough to generate a consistent number of LTR retrotransposon detections. The one exception being PE0075, which had an abnormally low number of LTR retrotransposons. In addition to this the tool is capable of removing any LTR retrotransposons without the requisite conserved domains for reverse transcriptase, RNase and integrase. One thing that can be gleaned from a comparison between the results of the first filter and the results of the second filter is that the number of intact LTR's does not necessarily correlate with a high number of active LTR retrotransposons. Before the LTR retrotransposons without the conserved domains were filtered out, maize lines EP1, F7 and W22 had the highest numbers of LTR retrotransposons. In a previous stage of this project, it was anticipated that these lines would have higher than average numbers of active LTR retrotransposons after applying the second filter (Wiggington,

2022). This was not the case. EP1 has the highest number of detections in Table 3, with four, but F7 and W22 did not have a high frequency of detections. Maize line F7 had one detection and W22 had no detections. It's possible that maize line F7 and W22 have high levels of LTR retrotransposons, but average levels of transposition, meaning the majority of the LTR retrotransposons in these two maize lines are old and inactive. The other possibility is that our logos are too stringent, and that the conserved domains we are looking for has more functional permutations than exists in any of the publicly available databases. However, this is less likely since proteins like reverse transcriptase are highly conserved (Sundaravaradan et al., 2005).

The 21 sequences that did not have all of the conserved domains typical of an active LTR retrotransposon had interesting similarities to each other. All of the LTR retrotransposons on chromosome five had very similar sequences. They had the same three conserved domains, none of them had less, and none of them had protease. They also had the same TSD sequence, and PPT. While more analysis would have to be performed, these are likely the same retrotransposon represented across many different maize lines. The tool developed picked out the same sequence in nine of sixteen maize lines. It shows that the tool could be made more restrictive than it is currently to search for just a single LTR retrotransposon across multiple large genomes. The six annotated sequences found by the tool - and found to have the desired domains - are likely capable of transposition. However, one of them NC358 Chromosome 1 does not have identical TSD's. While the TSD's are not required for transposition, it is a signifier that the LTR retrotransposons have picked up mutations through ectopic recombination (Jedlicka, 2020). This means that it is likely older than the other five typical sequences and has begun to degrade. It

would be interesting to cross this maize line with a line that is known to have no LTR retrotransposon to see if something that is still largely intact is still capable of transposition. Regardless, all six of these maize lines are good candidates for future crosses as they are far more likely to have offspring with transpositions than other maize lines.

Conclusion: The bioinformatic tool developed for this project successfully identified LTR retrotransposons with the domains and features needed for transposition. Out of over three million LTR’s initially identified, just twenty-seven were pulled from the database as active LTR retrotransposons. Even the unrefined iteration reduced this number to just 1,792 likely active LTR retrotransposons. This tool can be of use in future experiments and data analysis to significantly cut time spent crossing and growing maize. Plants with little likelihood to contain active LTR retrotransposons can be ignored, and significant time and money saved. In the future further refinement of the tool will make it more accurate and variations could be made for other transposable elements. The tool can also work with other plant genomes, since the domain and structure of LTR retrotransposons are similar in plants kingdoms.

REFERENCES

Abascal-Palacios, G., Jochem, L., Pla-Prats, C. *et al.* Structural basis of Ty3 retrotransposon integration at RNA Polymerase III-transcribed genes. *Nat Commun* **12**, 6992 (2021). https://doi.org/10.1038/s41467-021-27338-w

Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, et al. (2009) Exceptional Diversity, Non-Random Distribution, and Rapid Evolution of Retroelements in the B73 Maize Genome. PLOS Genetics 5(11): e1000732. https://doi.org/10.1371/journal.pgen.1000732

Boeke JD, Stoye JP. Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: Coffin JM, Hughes SH, Varmus HE, editors. Retroviruses. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, New York: 1997. pp. 343–435

Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 2009 Oct;10(10):691-703. doi: 10.1038/nrg2640. PMID: 19763152; PMCID: PMC2884099.

Crooks GE, Hon G, Chandonia JM, Brenner SE WebLogo: A sequence logo generator, *Genome Research*, 14:1188-1190, (2004)

Dooner, H. K., Wang, Q., Huang, J. T., Li, Y., He, L., Xiong, W., & Du, C. (2019). Spontaneous mutations in maize pollen are frequent in some lines and arise mainly from retrotranspositions and deletions. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(22), 10734–10743.

Eickbush TH, Jamburuthugoda VK. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res. 2008 Jun;134(1-2):221-34. doi: 10.1016/j.virusres.2007.12.010. Epub 2008 Feb 7. PMID: 18261821; PMCID: PMC2695964.

Finnegan DJ. Retrotransposons. Curr Biol. 2012 Jun 5;22(11):R432-7. doi: 10.1016/j.cub.2012.04.025. PMID: 22677280

Goodier, J. L., & Kazazian, J. H. H. (2008). Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. *Cell*, *135*(1), 23–35. https://doi-org.ezproxy.montclair.edu/10.1016/j.cell.2008.09.022

Hashimoto K, Jouhilahti EM, Töhönen V, Carninci P, Kere J, Katayama S. Embryonic LTR retrotransposons supply promoter modules to somatic tissues. Genome Res. 2021 Nov;31(11):1983-1993. doi: 10.1101/gr.275354.121. Epub 2021 Oct 21. PMID: 34675070; PMCID: PMC8559712.

Jedlicka Pavel, Lexa Matej, Kejnovsky Eduard. What Can Long Terminal Repeats Tell Us About the Age of LTR Retrotransposons, Gene Conversion and Ectopic Recombination? Frontiers in Plant Science. 2020; 11. Doi:10.3389/fpls.2020.00644

Kim TM, Hong SJ, Rhyu MG. Periodic Explosive Expansion of Human Retroelements Associated with the Evolution of the Hominoid Primate. J Korean Med Sci. 2004 Apr;19(2):177-185. https://doi.org/10.3346/jkms.2004.19.2.177

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020 Jan 8;48(D1):D265-D268. doi: 10.1093/nar/gkz991. PMID: 31777944; PMCID: PMC6943070.

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017 Jan 4;45(D1):D200-D203. doi: 10.1093/nar/gkw1129. Epub 2016 Nov 29. PMID: 27899674; PMCID: PMC5210587.

Mount SM, Rubin GM. Complete nucleotide sequence of the Drosophila transposable element copia: homology between copia and retroviral proteins. Mol Cell Biol. 1985 Jul;5(7):1630-8. doi: 10.1128/mcb.5.7.1630-1638.1985. PMID: 2410772; PMCID: PMC367281.

Potter CJ, Luo L. Splinkerette PCR for mapping transposable elements in Drosophila. PLoS One. 2010 Apr 13;5(4):e10168. doi: 10.1371/journal.pone.0010168. PMID: 20405015; PMCID: PMC2854151.

SanMiguel, P., Vitte, C. (2009). The LTR-Retrotransposons of Maize. In: Bennetzen, J.L., Hake, S. (eds) Handbook of Maize. Springer, New York, NY. https://doi.org/10.1007/978-0-387-77863-1_15

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022 Jan 7;50(D1):D20-D26. doi: 10.1093/nar/gkab1112. PMID: 34850941; PMCID: PMC8728269.

Sundaravaradan V, Hahn T, Ahmad N. Conservation of functional domains and limited heterogeneity of HIV-1 reverse transcriptase gene following vertical transmission. Retrovirology. 2005 May 26;2:36. doi: 10.1186/1742-4690-2-36. PMID: 15918905; PMCID: PMC1166575.

The UniProt Consortium, UniProt: the Universal Protein Knowledgebase in 2023, *Nucleic Acids Research*, Volume 51, Issue D1, 6 January 2023, Pages D523–D531, https://doi.org/10.1093/nar/gkac1052

Valencia, J.D., Girgis, H.Z. LtrDetector: A tool-suite for detecting long terminal repeat retrotransposons de-novo. *BMC Genomics* **20**, 450 (2019). https://doi.org/10.1186/s12864-019-5796-9

Wiggington Shay, Du Charles. (2022) Active LTR Retrotransposons Shaping the Dynamic Evolution of Maize Genomes. Montclair State University.

Woodhouse MR, Cannon EK, Portwood JL, Harper LC, Gardiner JM, Schaeffer ML, Andorf CM. (2021) A pan-genomic approach to genome databases using maize as a model system. BMC Plant Biol 21, 385. doi: https://doi.org/10.1186/s12870-021-03173-5.

APPENDIX A

Annotated LTR Retrotransposons

Legend:

TSD

Long Terminal Repeat

Primary Binding Site

Reverse Transcriptase

RNase

Integrase

Polypurine Tract

B97

>chr9:143627237-143631957

GTTTTTGTTAAAAGATAGAAGGATATATCCATCCTTGGGAGCTGCGATCTCCCTGATTTCCTGAACCAATC
TATTCTTGGGAAGTTATTATCTCCCTGATTTCTTGGAGCTACACCTCCGTGATTCTCTAGAGTTGTTACCAT
GTAAATCCTGTGTACCTCGTTCCTATATATTGAAGGATGCCGTGACCCCTAGAGGGCAGGCTGTGCATTG
CGTCAACATGGTATAGGAGCCAACTATCGATCCCAGATGTCCACAGACTCCTCTCTTTTTTCCGCCTCCTCC
CCTTCCACCATGACTACCAACGCTCCAGCATTCCAATCTGCCTGCGTTCCAATTCCTCTGCACCATGTGGTA
ACCATCCGTCTTACCAAGAGCAACTATCTTCTATGGCGCGCCCAACTTGTTCCGTTCCTCAGGAGCGCGAA
ACTCATGGGTTATCTTGATGGCACATTGCCGGCGCCTGCCCAACTGATCGCCTCTTCAACGGCGACTGGT
GCTGCGCAAGTGGCAAATCCAGCATATATACGCTGGTATGATCAAGATCAGCAGCTTCTCAGCGGACTTC
TCTCATCGATGACTGAAGATGTTCTTCGGGACGTGGTTGCAGCCACCTCATCCAAGGAGGTGTGGGACTC
TCTGCATAAGAAGTTCTCCTCCTCCACTCGAGCTCGCACGGTTCAGATCCGTGTTGAACTCGCTACGGCAA
AGAAGCGTGACTCAACAGCTGCTGACTACTATCGCAAGATTACAGGGCTGGCAAATGAGCTGGCTGTTG
CTGGCGCTCCACTCCAAGATGATGAGGTTCTCTCTTATTTACTTGCTGGCCTTCCTGCTGAATATGATCCGT
TTGTTACTTCCATTACCACCAAGACAGATCCTCTCTCGCTTGACGATGTGTTCGCCCATCTGATGGCGTTC
GAAGCACGACAACTCCAGCATCAGACTGAGATTCAACTAAACACCGACATGTCCGCGAACTATGTTGGG
CGTGGTGGTCGTGGTCATCGCGGTCGTGGCCGCAACACTCGTGGCCGCGGGCGTTTTGCACCTAGGGGG
GTGCGCCTCCTCGCGCATCCGGTGGTCATGGCTCATCTTCTCGTCCCACTTGTCAAATTTGTGAAAAGGAA
GGACATACAACAATCAAGTGCTGGTATCGCATGGATGAATCCTACACTGAGACACCTTCATCCTCAGCTC
TGGCGGCCACTACTTCGTACAAGATCGATCCAAACTGGTATCTTGATACGGGCGCTACCGATCACATAAC
CATCGATCTCGATCGTCTGGCCATACGAGAACGCTACAACGGTGGTGAGCAAGTCCACGTCGGTAATGG
AGCAGGTTTGCGCATTTTGCATATTGGCCATTCTTCCATTACTACTTCTATGCGTCCACTTGCATTGCGCAA
TATTCTTCATGTTCCACAAATTGCTAAGCATCTTCTTTCTGTTCATAAATTCTCTCTTGATAATGATGTATAT
TTTGAATATCATCCCTGGCATTTTTCCATTAAGGACCGGCACACCAAGAAAAGCCTGCTGGACGGGAAGT
GTGAGGCCGGCCTATATCCCATCAAGCCTTGTGATGTCGTCGCCCTTCATCAAGCTTTGATTAGTAGGACT
CCAAGTCGCGCCCAATGGCATGCTCGCCTTGGGCATCCCTCATCTCAAGTGGTCCAGTCAATTCTGCATCT
TAATAAAATTTTATGTCCTAGTGAGTCTTCCCTTCCAGTTTGTAATGCTTGTCAGTTAGCAAAAAGTCATCA

ATTACCCTATAATAATTCGGTTCATCAGTCGACCTCTCCTTTAGAGATGATTTTTTCCGATGTCTGGGGTCC
TGCGCCCCAATCTGTTAGTGGTTTCAAATATTATATCAGTTTCATCGATGATTTTAGTAAGTTTTGTTGGAT
CTATTTGATGCATGATCGTTCTGAAGTCCCTCGTATTTTTCGGCAGTTTCAAACTCATGTCGAACGTCTTTT
AAACACGAAAATTAAATGTGTTCAGTCTGACTGGGGAGGGGAATATCAGAAAATTCATAACACTCTTTTC
CGCCCGCTTGGCATTGCTCACCGTGTTTCTTGTCCACATACACATCAACAAAATGGTTCTGCTGAGCGTAA
ACATCGTCACATTGTTGAAACTGGCCTAGCTCTTCTTGCCCATGCACATGTTCCTCTAAAATTTTGGGATG
AGGCTTTTCTTACTGCTACGTATCTTATCAATCGTCTTCCTACTCGTGTCATAGACAACAAGTGCCCGTTAG
AGCGTCTGCTTAATACTCCTCCAAATTACTCCATGCTCAAAATTTTTGGATGTGCATGCTGGCCGCATCTTC
GAGCTTATAATAATCGCAAACTTGCTTTTCGCTCCAAAGAATGTGTTTTTCTTGGATACAGTTCGCTTCACA
AAGGGTATAAGTGCCTTGACATAGAATCTGGTCGCATCTATATTTCTCGGGATGTTGTTTTTGATGAAAAT
ATTTTTCCTTTTGCACGCTCAGCAGCTTCTCATCTTGTCCAATCCCATTCCGATTGTTCCTCCACATTGCGCA
AATTGCAGTCGGACATTAGTAGTACTGATGTGAATTATGATCATGTGCATATGTCGTTGCCTAATGCCTGC
TAACATTGTTCATGCAGAAAATCCAGCCACCTCATCCCTGCCTGAGTCTCAAGAATGGCTACAGGTTGCA
GCGCCAGCGCCTGCGCTAAATGCGACCACCTCTGCTACAAGAAGGCTCGGCGAATCTTCCTCCTTGGCTA
CATCTGGTCAGCCTGCTGATACTGGCACATTTACACCAACAGAACATGTGCCTGGGTCATCTGCTTCACCT
CAGTTGCCTTCCGCCGAGGTGCCTGCTGGGACCACCTTGCCTGCTGATCGTGTACTGCATCAGGATACTG
TTGCTCCTTCGCCTGAATCCCATCCAGATACACATCCATACGGCACCAGGTTGAAACACAACATCAGGATC
CCCAAGAAACGAACAGACGACACTGTTGCATGGTCAGTTGTTAAGTCCCCTACTGAACCTACAAATTATC
GTATTGCTATGGAAAATCCTTTTTGGCTCCAGGCTATGAATGATGAATTTCAGGCACTGTTGAAAAATAA
GACGTGGCACTTAATTCCCCCTAATCCAACTCTTAATCTTATTGATTGCAAATGGGTATTTCGGTTAAAGC
AGCTTCCTAATGGATCTATTGATCGCTATAAAGCTCGCCTGGTCGCCAAAGGTTTCACCCAGCAGTATGG
TGTGGATTATGATGAGACATTTAGTCCGGTTGTTAAGCCCACAACCATTCGTCTATTACTATCTCTTGCTAT
TTCTCGAGGATGGTCTATGCGTCAGATTGACATACAAAATGCTTTTCTTCATGGTATTCTTAATGAAGACG
TCTATATGAAGCAACCTCCTGGTTTTCGACTCCAAGCATCCGGACTATGTTTGCAAATTGGACAAAGCCTT
GTATGGCTTGAAGCAATCGCCCCGTGCGTGGTTTTCCCGTCTTAGCTCCAAATTGATACAATTGGGATTTA
CTGCTTCCAAGGCCGACGTATCTTTATTCATCTTTAATCAAGATGGTATTCAGATCTACATTCTTATCTACG
TTGATGATATTATTATTCTCAGTTCATCAACAGTGGCCACTGACAAACTTCTCGTTCAACTCCGGGATGAT
TTTGCAGTAAAAGATCTTGGACAGCTTAGCTATTTTCTTGGCATTGAGGTCCATCACACATCTGATGGTAT
TGTGTTGCTACAGCGCAAATATATTCGTGATCTATTGAGTCGCACCAACATGCTGTCTTCTAACGGTGTGC
CCACTCCTATGCTTCCCACTGAAAAATTATCTCTTGATGGTGGTGAAAAGCTTTCTTCAGATGATGCGACT
AAATATCGAAGTGCTGTGGGTGCCTTGCAATATCTTTTATTGACTCGCCCTGACATCTCCTTCAGTGTTAA
TCGGGTTTGCCAATTTATGTCTGCTCCAACTGATGTTCATTGGGGAGCTGTAAAACGGATTCTTCGTTATC
TTCATGCAACTATGGATTTCGGTCTCCGATTTAGTCGCAGGGGATCTTCTCTTCTCAGTGCTTTTTCGGATG
CTGATTGGGCTGGTAATCCGGATGATCGTCGTAGCACCGGTGGCTATGTTATATTTTTTGGTGGGAATCT
TGTTTCTTGGAGTTCACGGAAGCAAGCAACAGTTTCTCGCTCAAGCACTGAATCTGAATATAAAGCTGTT
GCTGATGCTACGGCTGAAGTTATATGGCTCCAAGTTTTATTGCGGGAACTTGGAGTTTCGCAACCACGTT
CTCCTACTCTCTGGTGTGACAATCTTGGTGCTACGTATCTATCGGCCAATCCGATTTTTCATCGTCGCTCCA
AGCATATTGAAGTTGATTATCACTGCATTGACAAAGCCATTGGCGGCACCTGCTTTTATTCGTTTTAGAAA
CAATCTGAACTTAGTATCTCTACGTCCAGATTGAGGGGGGATGTTAAAAGATAGAAGGATATATCCATCC
TTGGGAGCTGCGATCTCCCTGATTTCCTGAACCAATCTATTCTTGGGAAGTTATTATCTCCCTGATTTCTTG
GAGCTACACCTCCGTGATTCTCTAGAGTTGTTACCATGTAAATCCTGTGTACCTCGTTCCTATATATTGAA
GGATGCCGTGACCCCTAGAGGGCAGGCTGTGCATTGCGTCAACAGTTCT

CML103

>chr1:222665332-222669542

ACCTTATTCAAGTTCTTTGTAACACCCACTTTGTAAATTGCTAGAAATAATACCAATAGAGAAACAATTTCT
CTTTATGTGAGTTTGATCTGATCTAGGTGGCGTTTCTCAGTTGACATTTGCGCCGACGATCCTTAGTTCGT
TTTATATTCATTATTTTATTTTGTAATAAGTCTTCCGCTATGTAATAAATACTCTGATGTTTTATGACATTTA
TCTCTATACACTCTGTTATTATATATGTTGTCTTCTTTGGCGCATGTATGAGATGCACCTGGCTTTGTTCCTT
AAAGCCGGGTGTTAAAGAAGTGGTATCAGAGGAAATGTTGACTGTAGGACGAAACCTAGATAGAAATG
GACAAAACCCTTCCCTACTTACCTTACTCTGATTCATTCTATACTTATCTCATCTTGATCCTATCTCACCTTTT
GCTATTCCACTCTGATTACTCTCATCTTTTCTACTCTAAGACAAGATGGATTTCGCACCCTGAAATCCTATG
ATCTTTTTAAGAGATAGGGAACCTAAGACGAAAATAAAACTATTTTCTCTATTTTAAATGTTGGTTGATTG
TTCTGATGATCAATGTCTGATTTGCTTCTTTGATTGATTGAAATAGTATAGACGGGCATCTTAGCATGTAC
CACCATAAGGTTATGAGTTAGCTTTAGTAGGTGACACACTTAGCTAATAAATCCCCCAAAGTAACATGCTT
CGTAATTACATGCCTTGTCTACAACATTTTCTTTCGTACCCTATGTTCCCAACCCAGATGGGCTACCCCTAT
AGTGTTAGAAGAGAGCACTATAGACGTAATCCTTGGTATGTCATGGTTAAGAAAGGCAAAGGCAGTTTA
TGCACTGTGCTAACGGAACCGTAGAACTCACCAGTTCCAAAGGACAAAGATTTGAAGATGAAGTTGCAG
TAACTACCGTCACTTGACTAGCGGCATTCTTAGTAGATGGGAAGTTTGTTGGTGTCAACATCCGTGTGGT
TAGAGATTTTCCGGATGTCTTCCCAGAAGAGTTACCAGGGATGCCACCATATAGGGAAGTTGAGTTCGTC
ATAGATCTCTTACCTGGGACTGCCCCTACTTTCAAACGTCCATACAGGAGGTCTGTAGAAGAGTTAAAGG
GAGTGAAAAAGCAATTAACGGAATTACAAGAGACTGGGTACATTCGTCCGAGTTCCTCACCTTGGGGAG
CACTGGTTCTGTTTGTACAAAAGAAGGATAGATCACAACGGATGTGTGTGGACTATAGATCACTTAATGA
TGTTACTGTGAAGAACAAGTATCCCTTACCCCGTATTGAGGATTTGTTTGACCAGATGCGAGGAGTCAGG
GTATTCTCGAAGATTGACCTCCGATCGGGATACCATCAGATGAGGATTAGACCATCGGATATTCCTAAGA
CGGCTTTCTCGACTCGATATGGATTATATGAGTTCACTGTTATGTCATTTGGATTAACCAATGCACTAGCT
TATTTCATGAATTTGATGAATAATGTGTTTATGGAGTATTTGGACAGATTCGTCGTGGTGTTCATCGACGA
TATTCTTATCTATTCTAGGAGTGAAAGTGATCATGAACAACATCTAAGGTTGGTGCTACAGAAGCTACGA
GATAATCAACTCTACGCTAAGTTCAGTAAGTGCGAGTTTTGGATTGACAAGGTGTCATTCCTTGGTCATAT
TATTTCTAATGGAGGAATAGCAGTGGATCCTGCTAAAGTGAAGGAGATAATGGAGTGGAGAGTGCCCAC
TACAGTTACTGAGATTCGGAGTTTCCTGGGACTCGCAGGATATTATCGGAGATTCATTGAAGGGTTCTCT
AAAATTGCCAAGCCTATGACATCGCTTTTGGAAAAAGGAAAAGAATTTAAGTGGGATGAGAAATGCCAA
GAAAGCTTTGATCAATTGAAGGAGAGATTGATGTCACCTCCGGTGTTGGTAATGCCAGACCTACAGAAA
GGATTTGATATCTATTGTGATGCATGTGGCCAAGGATTGGGATGTGTGCTCATGCAAGAAGGTCATGTG
ATTGCCTATGCATCTCGTCAGTTGCGGAAACATGAATTGAACTACCCCACTCATGATTTGGAATTGGCAGC
AGTTGTGCATGCACTTAAGATTTGGAGACACTACATCATGGGAACTAAGTGCCAAGTATACACGGATCAT
AAAAGTTTGAAGTATATATTCACCCAGAAGGATCTCAACCTTAGACAACGCCGTTGGTTGGAGCTCATTA
AGGATTATGATTTGGAGATTCACTATCACCCAGGCAAGGCGAATTTGGTCGCAGATGCCTTGAGTCGGA
AGGAACATGTTCACTCCGCTTTCGTTGTCCAGCTACCCGATGAATTAGCAAAGGATTTTGAGAAGCTCAA
CCTGGGAATTGTTACTCACACTGAAGGAGTTACCATTGAGTTGGAAACTACCTTGGAGCAAGAGATTCGT
AAAGGTCAGGTTGGTGATGCTAAGATTCAAGAGATTAAGGATCTTATTACTGATGGTCGAGTTCCGGAAT
TCACGGAGGATGAGCAAGGTACTATATGGTTCAAAGATCGGATATGTGTTCCAGAGATCGATAGCCTCC
GGGAGACTATATTGAAGGAAGCCCATGATTCTGATTATTCAATCCACCCTAGAAGCACTAAGATGTATCA
AGATTTGAAGCAAAAGTATTGGTGGTATGGATTGAAAAGAGATGTGGCTGCTCATGTGGCTAAGTGTGA
TGTGTGTCAAAGAGTCAAGGCCGAACATCAAAGACCCGCTGGACTATTGCACCCATTGAAGATACCGGA
GTGGAAATGGGAAGAGATTGGCATAGACTTCATTGTTGGATTACCCCGCACGTCTGCTGGATATGATTCT
ATATGGGTGATTGTGGATAGGCTAACTAAGGTAGCCCACTTTATTCCAGTGAAGACCACATATTCAGGAG
CCAAGTTGGCAGAGTTGTATATGGCACGAATTGTATGCCTGCATGGTGTGCCAAAGAAAATCGTGTCTGA

TCGAGGATCACAGTTTACCTCTCGGTATTGGAAGAAGTTACACGAGTCAGTGGATACGAGATTGAATTTT
AGTACAGCTTATCACCCTTAGACTGATGGACAGACTGAAAGGACTAATCAAGTATTGGAAGATATGTTGA
GAGCATGTGCCCTTCAGCATGGAGGAAGTTGGGATAAAAGTTTGCCTTATGCTGGGTTCTCATATAATAA
TAGTTATCAGGCCAGTCTGAAGATGTCACCTTTTGAAGCTTTATACGGGAGAAAGTGCAGAACTCCTTTG
TATTGGGATCAGACTGGAGAAAGACAGATCTTTGGACCTGAACTTATTCAAGAGGCAGAAGAACAAGTT
TGTATAATTCGAGAGAATTTGAGGGTAGCTCAAACCAGGCAAAAGAGTTATGCTGATAATAGAAGAAGA
CCACTGGAATTTGAAGAAGGAGATTATGTGTATCTCAAGGTGTCACCACTTCGTGGAATGAGGAGATTCA
AAGTTAAGGGCAAATTATCCCCTCGCTATATTGGACCATTCTTGATCTTTAGGCGAGTTGGAGAGATGGC
ATACCAACTGGAGTTACCTGCTAGTCTATCGGATGTGCATAATGTGTTCCATGTATCTCAACTGAAGGGG
TGTCTCCGTGTCCCTGAGGAACAGTTACCAATGGAAGAGCTCAGTGTTCAGGGTGATTTGACTTACACGG
AATACCCGATCAAGATTTTGGATACATTGACTCGAGTTACAAGAAATAAGGTGATAAAGATGTGCAAAGT
ACAATGGAGTCACCATGGTGAAGACGAAGCGACATGGGAAAGGGAAGAAGAGCTTCACATAGATTTTC
CCCATCTTTTCTCTAGTTCTTCCTAAATCTCGAGGACGAGATTATTTTTAAGGGGGGTAGGATTTGTAACA
CCCACTTTGTAAATTGCTAGAAATAATACCAATAGAGAAACAATTTCTCTTTATGTGAGTTTGATCTGATCT
AGGTGGCGTTTCTCAGTTGACATTTGCGCCGACGATCCTTAGTTCGTTTTATATTCATTATTTATTTGTA
ATAAGTCTTCCGCTATGTAATAAATACTCTGATGTTTTATGACATTTATCTCTATACACTCTGTTATTATATA
TGTTGTCTTCTTTGGCGCATGTATGAGATGCACCTGGCTTTGTTCCTTAAAGCCGGGTGTTACATTCTTCCA
CCT

>chr9:143942027-143946747

GTTTTTGTTAAAAGATAGAAGGATATATCCATCCTTGGGAGCTGCGATCTCCCTGATTTCCTGAACCAATC
TATTCTTGGGAAGTTATTATCTCCCTGATTTCTTGGAGCTACACCTCCGTGATTCTCTAGAGTTGTTACCAT
GTAAATCCTGTGTACCTCGTTCCTATATATTGAAGGATGCCGTGACCCCTAGAGGGCAGGCTGTGCATTG
CGTCAACATGGTATAGGAGCCAACTATCGATCCCAGATGTCCACAGACTCCTCTCTTTTTTCCGCCTCCTCC
CCTTCCACCATGACTACCAACGCTCCAGCATTCCAATCTGCCTGCGTTCCAATTCCTCTGCACCATGTGGTA
ACCATCCGTCTTACCAAGAGCAACTATCTTCTATGGCGCGCCCAACTTGTTCCGTTCCTCAGGAGCGCGAA
ACTCATGGGTTATCTTGATGGCACATTGCCGGCGCCTGCCCAACTGATCGCCTCTTCAACGGCGACTGGT
GCTGCGCAAGTGGCAAATCCAGCATATATACGCTGGTATGATCAAGATCAGCAGCTTCTCAGCGGACTTC
TCTCATCGATGACTGAAGATGTTCTTCGGGACGTGGTTGCAGCCACCTCATCCAAGGAGGTGTGGGACTC
TCTGCATAAGAAGTTCTCCTCCTCCACTCGAGCTCGCACGGTTCAGATCCGTGTTGAACTCGCTACGGCAA
AGAAGCGTGACTCAACAGCTGCTGACTACTATCGCAAGATTACAGGGCTGGCAAATGAGCTGGCTGTTG
CTGGCGCTCCACTCCAAGATGATGAGGTTCTCTCTTATTTACTTGCTGGCCTTCCTGCTGAATATGATCCGT
TTGTTACTTCCATTACCACCAAGACAGATCCTCTCTCGCTTGACGATGTGTTCGCCCATCTGATGGCGTTC
GAAGCACGACAACTCCAGCATCAGACTGAGATTCAACTAAACACCGACATGTCCGCGAACTATGTTGGG
CGTGGTGGTCGTGGTCATCGCGGTCGTGGCCGCAACACTCGTGGCCGCGGGCGTTTTGCACCTAGGGGG
GTGCGCCTCCTCGCGCATCCGGTGGTCATGGCTCATCTTCTCGTCCCACTTGTCAAATTTGTGAAAAGGAA
GGACATACAACAATCAAGTGCTGGTATCGCATGGATGAATCCTACACTGAGACACCTTCATCCTCAGCTC
TGGCGGCCACTACTTCGTACAAGATCGATCCAAACTGGTATCTTGATACGGGCGCTACCGATCACATAAC
CAGCGATCTCGATCGTCTGGCCATACGAGAACGCTACAACGGTGGTGAGCAAGTCCACGTCGGTAATGG
AGCAGGTTTGCGCATTTTGCATATTGGCCATTCTTCCATTACTACTTCTATGCGTCCACTTGCATTGCGCAA
TATTCTTCATGTTCCACAAATTGCTAAGCATCTTCTTTCTGTTCATAAATTCTCTCTTGATAATGATGTATAT
TTTGAATATCATCCCTGGCATTTTTCCATTAAGGACCGGCACACCAAGAAAAGCCTGCTGGACGGGAAGT
GTGAGGCCGGCCTATATCCCATCAAGCCTTGTGATGTCGTCGCCCTTCATCAAGCTTTGATTAGTAGGACT

CCAAGTCGCGCCCAATGGCATGCTCGCCTTGGGCATCCCTCATCTCAAGTGGTCCAGTCAATTCTGCATCT
TAATAAAATTTTATGTCCTAGTGAGTCTTCCCTTCCAGTTTGTAATGCTTGTCAGTTAGCAAAAAGTCATCA
ATTACCCTATAATAATTCGGTTCATCAGTCGACCTCTCCTTTAGAGATGATTTTTTCCGATGTCTGGGGTCC
TGCGCCCCAATCTGTTAGTGGTTTCAAATATTATATCAGTTTCATCGATGATTTTAGTAAGTTTTGTTGGAT
CTATTTGATGCATGATCGTTCTGAAGTCCCTCGTATTTTTCGGCAGTTTCAAACTCATGTCGAACGTCTTTT
AAACACGAAAATTAAATGTGTTCAGTCTGACTGGGGAGGGGAATATCAGAAAATTCATAACACTCTTTTC
CGCCCGCTTGGCATTGCTCACCGTGTTTCTTGTCCACATACACATCAACAAAATGGTTCTGCTGAGCGTAA
ACATCGTCACATTGTTGAAACTGGCCTAGCTCTTCTTGCCCATGCACATGTTCCTCTAAAATTTTGGGATG
AGGCTTTTCTTACTGCTACGTATCTTATCAATCGTCTTCCTACTCGTGTCATAGACAACAAGTGCCCGTTAG
AGCGTCTGCTTAATACTCCTCCAAATTACTCCATGCTCAAAATTTTTGGATGTGCATGCTGGCCGCATCTTC
GAGCTTATAATAATCGCAAACTTGCTTTTCGCTCCAAAGAATGTGTTTTTCTTGGATACAGTTCGCTTCACA
AAGGGTATAAGTGCCTTGACATAGAATCTGGTCGCATCTATATTTCTCGGGATGTTGTTTTTGATGAAAAT
ATTTTTCCTTTTGCACGCTCAGCAGCTTCTCATCTTGTCCAATCCCATTCCGATTGTTCCTCCACATTGCGCA
AATTGCAGTCGGACATTAGTAGTACTGATGTGAATTATGATCATGTGCATATGTCGTTGCCTAATGCCTGC
TAACATTGTTCATGCAGAAAATCCAGCCACCTCATCCCTGCCTGAGTCTCAAGAATGGCTACAGGTTGCA
GCGCCAGCGCCTGCGCTAAATGCGACCACCTCTGCTACAAGAAGGCTCGGCGAATCTTCCTCCTTGGCTA
CATCTGGTCAGCCTGCTGATACTGGCACATTTACACCAACAGAACATGTGCCTGGGTCATCTGCTTCACCT
CAGTTGCCTTCCGCCGAGGTGCCTGCTGGGACCACCTTGCCTGCTGATCGTGTACTGCATCAGGATACTG
TTGCTCCTTCGCCTGAATCCCATCCAGATACACATCCATACGGCACCAGGTTGAAACACAACATCAGGATC
CCCAAGAAACGAACAGACGACACTGTTGCATGGTCAGTTGTTAAGTCCCCTACTGAACCTACAAATTATC
GTATTGCTATGGAAAATCCTTTTTGGCTCCAGGCTATGAATGATGAATTTCAGGCACTGTTGAAAAATAA
GACGTGGCACTTAATTCCCCCTAATCCAACTCTTAATCTTATTGATTGCAAATGGGTATTTCGGTTAAAGC
AGCTTCCTAATGGATCTATTGATCGCTATAAAGCTCGCCTGGTCGCCAAAGGTTTCACCCAGCAGTATGG
TGTGGATTATGATGAGACATTTAGTCCGGTTGTTAAGCCCACAACCATTCGTCTATTACTATCTCTTGCTAT
TTCTCGAGGATGGTCTATGCGTCAGATTGACATACAAAATGCTTTTCTTCATGGTATTCTTAATGAAGACG
TCTATATGAAGCAACCTCCTGGTTTTCGACTCCAAGCATCCGGACTATGTTTGCAAATTGGACAAAGCCTT
GTATGGCTTGAAGCAATCGCCCCGTGCGTGGTTTTCCCGTCTTAGCTCCAAATTGATACAATTGGGATTTA
CTGCTTCCAAGGCCGACGTATCTTTATTCATCTTTAATCAAGATGGTATTCAGATCTACATTCTTATCTACG
TTGATGATATTATTATTCTCAGTTCATCAACAGTGGCCACTGACAAACTTCTCGTTCAACTCCGGGATGAT
TTTGCAGTAAAAGATCTTGGACAGCTTAGCTATTTTCTTGGCATTGAGGTCCATCACACATCTGATGGTAT
TGTGTTGCTACAGCGCAAATATATTCGTGATCTATTGAGTCGCACCAACATGCTGTCTTCTAACGGTGTGC
CCACTCCTATGCTTCCCACTGAAAAATTATCTCTTGATGGTGGTGAAAAGCTTTCTTCAGATGATGCGACT
AAATATCGAAGTGCTGTGGGTGCCTTGCAATATCTTTTATTGACTCGCCCTGACATCTCCTTCAGTGTTAA
TCGGGTTTGCCAATTTATGTCTGCTCCAACTGATGTTCATTGGGGAGCTGTAAAACGGATTCTTCGTTATC
TTCATGCAACTATGGATTTCGGTCTCCGATTTAGTCGCAGGGGATCTTCTCTTCTCAGTGCTTTTTCGGATG
CTGATTGGGCTGGTAATCCGGATGATCGTCGTAGCACCGGTGGCTATGTTATATTTTTTGGTGGGAATCT
TGTTTCTTGGAGTTCACGGAAGCAAGCAACAGTTTCTCGCTCAAGCACTGAATCTGAATATAAAGCTGTT
GCTGATGCTACGGCTGAAGTTATATGGCTCCAAGTTTTATTGCGGGAACTTGGAGTTTCGCAACCACGTT
CTCCTACTCTCTGGTGTGACAATCTTGGTGCTACGTATCTATCGGCCAATCCGATTTTTCATCGTCGCTCCA
AGCATATTGAAGTTGATTATCACTGCATTGACAAAGCCATTGGCGGCACCTGCTTTTATTCGTTTTAGAAA
CAATCTGAACTTAGTATCTCTACGTCCAGATTGAGGGGGGATGTTAAAAGATAGAAGGATATATCCATCC
TTGGGAGCTGCGATCTCCCTGATTTCCTGAACCAATCTATTCTTGGGAAGTTATTATCTCCCTGATTTCTTG
GAGCTACACCTCCGTGATTCTCTAGAGTTGTTACCATGTAAATCCTGTGTACCTCGTTCCTATATATTGAA
GGATGCCGTGACCCCTAGAGGGCAGGCTGTGCATTGCGTCAACAGTTCT

>chr5:122728891-122732807

TGGTTCGATTGCTCCATGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTTGCGTGGAGT
TGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATAGTGGGCCGG
CCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAGAAACAGAAT
CCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGGCGCGGCTCA
CGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCATCTGGTATCGGATTCGGGGTCGGCTGCGT
CTTCATCCTCTTCCACGTGCTGTTCCTCGTCTTTGGCGGTTCGTGGGCGACTTCTTCTTCTGCTGGGTCCAT
CCAGTTCATCACTACGGCGCCCGCAACCAAACCATCGGTACAGACTCGTTTGGTGTTACAGGCTATGGCG
GATGGGGAGCGAGTTAGTCAGGAGAATTGGAGTCGTGTTATTGAGACGTTGGAGGGAGTGAGTAGCAG
GTTGAAGGATGTGGAGAGGGTGCAGCAACAAATCATGATGCGGACTGAATTAGCGGCATCAGTTGCTG
AACAGGCTGCGAAGGAGAGAAATGATTTGGGTCGTAAATTGGAGGAGACTGGAAGGGGAATGATATTT
GGGCTTTAGTGGTGTTGGAAGAGGTGTCTGCAGATCAGGGAGTACAGCAGCAGGGGGATCTTCAGCAG
TTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTACCTCCTGCTAGGGTGTATGATCATCA
TATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTATAGGTACTCTCCTCTTCACAAGGATG
AAATTGAGAGACAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATTGTTCCTAGTGTAAGCCCGTTTGCTTCT
CCGGTATTGTTGGTTAAGAAGAAAGATAGCACTTGGAGGTTTTGTGTGGACTACAGGAAATTAAATTCTT
TGACCATTAAGAACAGATTTCCTATGCCGCTTGTAGAAGAGATTCTGGATGAACTGGCTGGTACTAAATA
TTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAATGGGGGAAGGTGATGAGCATAAAACA
GCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCCTTTTGGGTTAACCAATGCTCCTGCAAC
CTTTCAATGTGCTATGAATTCTGTTTTGGCTCCTTTTCTTCGGAAATTTGTCATGGTTTTCATTGATGACATT
TTGGTCTATAGTGCTACATGGGAGGAACACTTGGCTCATATTAAGATGGTGCTGCAGACTTTGAGAGAG
CACCAGTTCTATCTGAAGAGGAGTAAATGTGTGTTTGGGAGGAATGAGCTCATTTATTTGGGACATATTA
TCTCTCACACAGGAGTGGCCACTGATCCTGCTAAGACAACAGCCATGGAGAAATGGCCTGTTCCTCAGTC
AGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTATAGGAGATTTGTGAGACATTATGGGAT
TATTGCAAAACCTCTTACTAATCTGTTGAAGCATAAGGCTTTTCAGTGGAGTCCTGAGGCTGATCAGGCTT
TTCAAGCGCTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCCCTACCCAGGTTTGATGAGTTATTTGTT
GTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGATGCAATGTGATAGACCAGTGGCATAT
CTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTATCCATTTATGAAAAAGAGTTTTTGGCTCTAATCA
TTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATTTGTGATTAAAACTGATCATCAGGCTT
TGAGTTTTCTGGATGATCAGCTTTTGCAATCTGATATGCAGAGGAAGGCTATGGCTAAGTTGATGGGGCT
TCATTTCAAAATTGTTTACAAAAAAGGCAAGGACAATGTGGCTGCCGATGCTCTATCCAGGATTGGTCAT
GTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGGGTGCAAGAACTCCTGAATGCTTATGTG
ACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGGCCAGTCCTAATGAGCAAGGTTTTTCTT
TACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGCTAACAATTCTGCCCTTCAAACTAAGTT
GATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGTGGTGAGGCCACTTACCAAAGGATCAA
GAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAGTTTGTGCAGCAATGTGAGGTGTGTCA
GCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAACCTTTACCAGTGCCTCAAGGGGCTTGG
CAGGATCTTACATTGGATTTCATTGAAGGTCTACCTAGATCTGAGGGATATGATACTATTTTGGTCGTGGT
GGATAGATATTCCAAATATGCTCATTTTTTTCCCCTCAAGCATCCTTTCACTGCTGTTGGAGTGGCACAGG
TGTTTCTTGACAATATTGTCAAGTTGCATGGGACACCAAAGTCTTTGGTTTCCGATAGAGACGAAGTGTTC
ACTAGTTTGTTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCAAATTGGCTCTCACTACTGCATATCACCCT
CAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGAAATGTATCTTCGCTGTTCTATTCATGAGG

CACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAATTTTGGTACAATACTTCATATCACACAGCTTTG
GGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTGTTTGCTGCTGCTCCTATGCTTTCTGGGGT
TGCTCCTAGTTCTGTGACTGAGGTGTTGGCTGATAGACAGTTACATGCTAGTATGCTGAAGGAGAAGTTG
GCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCATAGATCTGACAGAGAGTTTCAGATTGGA
GATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCGGTGGTTAGTCGACCCTGTCCTAAATTGG
CCTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGGGATGGTTGCTTATAAGTTGGATTTACCA
GTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGTTGAAGGCTTTTACTCCCAATTATTCTCCGGTG
TTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGATTATGCCATTGGCTGTGGTGGAAAGGAG
GATGGTGAAGAAGGGTAATCATCCTGTGGTGCAAGTAAGGGTGACTTGGTCCAATATGCCAGCTGAGGT
TACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTTCCTACAGCTCTTGATTGGGGACAATCAAGC
TCTTCCGGGGGGGGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGT
TTTGCGTGGAGTTGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAA
TAGTGGGCCGGCCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAAT
AGAAACAGAATCCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTT
GGCGCGGCTCACGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCACTCCAGTGATGATGCATG
GT

CML322

>chr3:200615039-200619992

AGTACTGTCAGGACCATGGTTAGTCTATCATCTTCAGTAGAGTATCTGAATATTCGTGGTTGGGCCGTCA
GCCCATGTAAGTACTAAGAATTGGCCGGCAGTATATAAGCCGGGGTGGTGGAGGGTTTAAAAACAAGG
AAAAGAACAGAACTCATTTACCTGTATCTTCTTCGTTGCCTCACGGACTCAGCCGTGTTCTCTGTCTTCCTC
TCGCTCTCCATCTTCCTGGTGTCTGAATTCCGTTCTCAAGCAAGTATTCAAGGGAAGAGGGTTGGGTTCGT
GACATTTGGTATCAGAGGATAGGTTCCCAAATTTTTTTTGGCGAGTGGGATCGTTCTGGCGCCGAAGAAG
AGACAAACCAAGTTGGAACTCCGAGTCGATTCGGCCATGGAGGACTTCGCGAAGAAGTTGGAGGGATT
GGAAGCTATGTTGCAGAAGACCTTCGATGCTGTGAGCGGCTTGGAGTCGTGGCGAACTAGCGCAGACCA
GTCGCTCGACACATTGCTTCAGAGAACGGATGAGGCGGGAACACGTCTGCAGCAATTGGAATCTGCACC
GATGCGACCTCCCCCGCCACCACCACCTCCGCCGCCTCACCGACCAGCTTCGCCAAAGTATTGCGATCCGT
TCAACATGCAGCAATTTCGACCACAAGCAGGCGGGTCGAGCACGCTGGACCCGGGTCTGGCTCCATTGG
GATCGTCGCGCCCACCTATGACAGGCGTGGAGAACGCTCATGGGCACCGTCTGCAGCAAACTCTCCGGG
ATGTTGGTGGTGGGATTTTAGGATCCCATCCACCACGCCCGGTCACGGGTATGAATCTGTCCCATGATCT
CCCAGATTTTGATGCACCTATGGGTTCTCGAGATCATTCGACTTTAATTCCGAAAATGGAGTTTCCTAAAT
TTGAGGAACTTAATCTGGTGAAGACTCGGGCAATGACTAAGGAGTACAATAGAGGTCATTATCGCACAG
GCAAGGAAAATGACAAACCAAAGGATGACAGTATGGAGGTGAACAAAATTAAAGTTGACCGAACGGAG
AGAAGTGACAAGTTAGAAGCATTGCGAAATTACAGAAGGAAAAATGGATTGTGTTTCAAGTGTGGAGA
GAAATGGGGTCAACAGCATAAGTGCCCTACCCAGATCTCGTTACATGTATTAGAAGAAATCTTGGATGCC
ATTGAATTGGACTATGAACAATTCAATAGGGAAGATAATGAAACGGTGGTGGAAGAACAAGTGTTAGCA
GTCCAAGAAACGGTATCTGGGAGTAAGAAAAGAAGTACACTACGATTACTGGGCATCATTGGCACGCAA
CAAGTCTTGATCTTGGTAGACTCGGGGAGTATTGGCACTTTTGTCAGTAAGGGATTAGTAGATCGTTTGA

TGTTACCAACACAGAGTTGTGCCCAGACACAGTTTCGGGCTGCTGATGGAGGAATTATGAGCTGTACTGA
ATGGGTGCCAAGGCTGCAATGGTTTATTCAGGGCTATTCGTTTGTTTCTCAGGCCAAGGTGTTACCATTAC
AATGTTATGACATGATATTAGGCGAAGATTGGCTTGAGGAAGTAAGTCCATTGTGGGTGGATTACAAGT
CAAAGCAACTAAGGTTTACCTATCAAGGTACTAGAGTGTCCTTGCAAGGGGTTCAAGATAACATCACGAG
CTGTCAGAGTATCTCAGGACATAAGTTCCAGGGATTACTTAATCATGGAGTTGTCTATCATTGCATACAGT
TGCTTCTGATAAATCAGGACAATGATTTGCCTACACAAGGTCCTGTCTCTCGAGTATGTGAGATTCAGAAT
ATTGAGCAACAGCCAATACCTCTCAGAATTCAGCAGCTCATTGATTCTTTTGCTCACCTCTTCAGTTCACCC
ACAGGACTTCCACCACATCGCTTAGCTGATCATCAAATTCCTCTAATGCCTGGTGCTCAACCGGTGAAAGT
GAGATCTTACAGATACTCTCCTGCTCAAAAATCTGAAATCGAGAAGCAATTGAAGGAAATGTTGACTCAA
GGGATAATTCGTCCGAGCACTAGCCCCTTTGCTTCTCCGGTAATTTTGGTCAAAAAGAAAGATGGGAGTT
GGCGATTTTGTGTTGATTATCGACACTTAAATGCAATCACAATCAAGAATAAACATCCCATGCCAATAGTG
GATGAGCTGCTAGATGAGTTGGCCGGAGCTTGTTGGTTTACTAAATTGGATTTTAGCGCTGGATACCATC
AGGTCTGCATGGCTGTTGGAGAGGAGTACAAGACAGCATTTAGAACTCATAATGGCCTGTATGAATTCAT
GGTTATGCCTTTTGGCCTTACAAATGCTCCTGCAACATTTCAAGGACTTATGAATCACATTTTTGCTGATTT
ACTAAGGCATGGAGTGCTGGTATTCATGGATGACATTTTGATTTACAGCCACACATTACATGAGCATGAG
CTGTTGTTGCAACAGGTGTTTCAGACTTTGAATCAACACAAGTTCTTGATTAAAAGAACTAAATGCTCTTT
TGCTCAACCATCCATAGAATACTTAGGCCATGTGATTTCAGCCCAAGGAGTATCTACTGAACCCTCCAAAG
TGGCCGCAGTCAAGAACTGGCCTATTCCCAAGAATGTGAAGCAACTCCGAGGTTTTTTAGGTCTCACAGG
ATATTATAGGAAATTCATCAAAGACTATGGAATTGTTAGCCGACCACTAACCTTATTGCTTAAGAAGGGTT
CAGTGTTCCGGTGGTCTGAGGAACAGCAAGTGGCTTTCACATGCTTAAACAAAAGATGGTTGAAGCTCC
AGTGTTGATGGTTCCGGATTTTTCTAAACAGTTTGTTCTGGAAACAGATGCTTGTGATCTGGGTATAGGT
GCAGTTTTGATGCAGAGTGAGCATCCTGTGGCATATCTGAGTAAGGCATTGGGAATGAAGAATCAGGCG
TTGTCAGTATATGAGAAAGAATGCATGGCTATATTGTTGGCAGTGGAAAAATGGAGGCCCTACTTACAAT
ATCAACCGTTTGTCATCAGAACTGATCATCACAGCTTGCGGTTCTTAACTGACCAACGGGTTTCCACCCGG
TTGCAACATAAGGCGTTAGTTAAATTGATGGACCTTCAATATACCATTCAGTACAAAAAAGGATCCACAA
ATGCAGCTGCTGATGCCTTGTCGCGTTGTCCTGACCCTCCAGACATTTATGCAGTTCCGGAAGTGGTTCCC
ACGTGGTTGCAAAAACTGCAGCAAGGCTATGATGATGATGAGGAAACTAAGAAACTTATTACAGCTGTT
ACCTTGGATCCAGACAGTCATAAGAATTATACATTAACTTCGGGTATTCTGAGGTTTAAAGGCCGTGTGT
GGGTGGGGAATAACGCATTGGCTCAGCAAAACATTTTACAGGCACTTCATTCAAGTGGGATTGGAGGTC
ATTCTGGTATTCAAGGTACCTATCATCGTGTGAAGCAACTTTTTGCGTGGCCTGCTCTAAAAAAATCAGTG
CAACAATATGTCCAATCTTGTGAGATTTGTCAGCAAGCTAAAACAGAACATGTGAAATACCCGGGATTAT
TACAGCCACTGCCCATTCCTCAACAAGCTTGGGAGATAGTGAGTTTGGATTTTATTGAAGGCTTGCCATCC
TCAGAAAATTATAATACTATCCTTGTGGTTGTGGATAAATTTACCAAATATGGACACTTCTTAGCCTTATCT
CATCCATTCACAACATTGAAGGTAGCTCAACTATTTTTCTCGCAAATTTACAAGTTGCATGGGCTGCCTCG
CATGATTATCTCGGATAGGGACAGGATCTTCACAAGTACGCTCTGGCAGGAACTGTTTCGACTGTCAGAT
ACAGTGCTAATGATGAGCTCGAGTTATCATCCACAGACAGATGGACAAACTGAGAGATTAAATCAGTGTT
TAGAGGGTTTCCTAAGGTGCACAGTGCATTCGTGTCCTCGACGATGGCATAAGTGGCTTCCCGTGGCAG
AATACTGGTATAATACTACCATTCACTCTTCCTTGGGTCGTTCGCCTTTTGAAGTTCTGTATGGGTATGTTC
CTCGCACATTAGGAATTGCTAATCTAAAGCTCTGTACTGTTCCTGATCTGGAAACATGGATGCAAGAAAG
AGAATTGTTGTCTGAATTGGTTAAGCAGCAGTTAGCACGTGCTCAACAACGAATGAAGAGCCAGGCAGA
TAAAAATCGGTCTGAGCGGAGTTTTGAAGTTGGTGATCGAGTATATTTGAAGCTTCAACCACATGTTCAA
TCATCGGTAGCTCTGCGATCTAACCAGAAACTTGCATTCAGATATTATGGACCCTACACTGTTCTTCAGAA
AGTAGGACAGGTCGCTTACAAGTTGGACCTTCCAGACAGCTCAAAAATACATCCAGTTGTCCATGTTTCCT
TATTGAAGAAATATATTTCTCCTCCACATGACATTTCTGCTGATCTTTCTTCTGTGTGCACGGATCCGTTCC

AGATATTGACGTCGGACACCATATTACAGCAACGGTGCATTCGTAAAGGTCGTGCTTTGGTTCAGCAGGT
CTTGGTTAAATGGGTGGGACTGCCGAATCATTTGGCTACCTGGGAGCATGTTGATGATATTAAGTGTAG
GTTTCCAGCAACAACAGCATGAGGGCATGCTGCATGTAAGGGAGGGGGAATGTCAGGACCATGGTTAG
TCTATCATCTTCAGTAGAGTATCTGAATATTCGTGGTTGGGCCGTCAGCCCATGTAAGTACTAAGAATTG
GCCGGCAGTATATAAGCCGGGGTGGTGGAGGGTTTAAAAACAAGGAAAAGAACAGAACTCATTTACCT
GTATCTTCTTCGTTGCCTCACGGACTCAGCCGTGTTCTCTGTCTTCCTCTCGCTCTCCATCTTCCTGGTGTCT
GAATTCCGTTCTCAAGCAAGTATTCAAGGGAAGAGGGTTGGGTTCGTGACAAGTAC

>chr5:120020951-120024869

TGGTTCGATTGCTCCATGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTTGCGTGGAGT
TGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATAGTGGGCCGG
CCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAGAAACAGAAT
CCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGGCGCGGCTCA
CGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCATCTGGTATCGGATTCGGGGTCGGCTGCGT
CTTCATCCTCTTCCACGTGCTGTTCCTCGTCTTTGGCGGTTCGTGGGCGACTTCTTCTTCTGCTGGGTCCAT
CCAGTTCATCACTACGGCGCCCGCAACCAAACCATCGGTACAGACTCGTTTGGTGTTACAGGCTATGGCG
GATGGGGAGCGAGTTAGTCAGGAGAATTGGAGTCGTGTTATTGAGACGTTGGAGGGAGTGAGTAGCAG
GTTGAAGGATGTGGAGAGGGTGCAGCAACAAATCATGATGCGGACTGAATTAGCGGCATCAGTTGCTG
AACAGGCTGCGAAGGAGAGAAATGATTTGGGTCGTAAATTGGAGGAGACTGGAAGGGGAATGATATTT
GGGCTTTAGTGGTGTTGGAAGAGGTGTCTGCAGATCAGGGAGTACAGCAGCAGGGGGATCTTCAGCAG
TTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTACCTCCTGCTAGGGTGTATGATCATCA
TATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTATAGGTACTCTCCTCTTCACAAGGATG
AAATTGAGAGGCAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATTGTTCCTAGTGTAAGCCCGTTTGCTTCT
CCGGTATTGTTGGTTAAGAAGAAAGATAGCACTTGGAGGTTTTGTGTGGACTACAGGAAATTAAATTCTT
TGATCATTAAGAACAGATTTCCTATGCCGCTTGTAGAAGAGATTCTGGATGAACTGGCTGGTACTAAATA
TTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAATGGGGGAAGGTGATGAGCATAAAACA
GCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCCTTTTGGGTTAACCAATGCTCCTGCAAC
CTTTCAATGTGCTATGAATTCTGTTTTGGCTCCTTTTCTTCGGAAATTTGTCATGGTTTTCATTGATGACATT
TTGGTCTATAGTGCTACATGGGAGGAACACTTGGCTCATATTAAGATGGTGCTGCAGACTTTGAGAGAG
CACCAGTTCTATCTGAAGAGGAGTAAATGTGTGTTTGGGAGGAATGAGCTCATTTATTTGGGACATATTA
TCTCTCACACAGGAGTGGCCACTGATCCTGCTAAGACAACAGCCATGGAGAAATGGCCTGTTCCTCAGTC
AGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTATAGGAGATTTGTGAGACATTATGGGAT
TATTGCAAAACCTCTTACTAATCTGTTGAAGCATAAGGCTTTTCAGTGGAGTCCTGAGGCTGATCAGGCTT
TTCAAGCGCTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCCCTACCCAGGTTTGATGAGTTATTTGTT
GTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGATGCAATGTGATAGACCAGTGGCATAT
CTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTATCCATTTATGAAAAAGAGTTTTTGGCTCTAATCA
TTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATTTGTGATTAAAACTGATCATCAGGCTT
TGAGTTTTCTGGATGATCAGCTTTTGCAATCTGATATGCAGAGGAAGGCTATGGCTAAGTTGATGGGGCT
TCATTTCAAAATTGTTTACAAAAAAGGCAAGGACAATGTGGCTGCCGATGCTCTATCCAGGATTGGTCAT
GTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGGGTGCAAGAACTCCTGAATGCTTATGTG
ACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGGCCAGTCCTAATGAGCAAGGTTTTTCTT
TACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGCTAACAATTCTGCCCTTCAAACTAAGTT
GATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGTGGTGAGGCCACTTACCAAAGGATCAA

GAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAGTTTGTGCAGCAATGTGAGGTGTGTCA
GCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAACCTTTACCAGTGCCTCAAGGGGCTTGG
CAGGATCTTACATTGGATTTCATTGAAGGTCTACCTAGATCTGAGGGATATGATACTATTTTGGTCGTGGT
GGATAGATATTCCAAATATGCTCATTTTTTTCCCCTCAAGCATCCTTTCACTGCTGTTGGAGTGGCACAGG
TGTTTCTTGACAATATTGTCAAGTTGCATGGGACACCAAAGTCTTTGGTTTCCGATAGAGACGAAGTGTTC
ACTAGTTTGTTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCAAATTGGCTCTCACTACTGCATATCACCCT
CAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGAAATGTATCTTCGCTGTTCTATTCATGAGG
CACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAATTTTGGTACAATACTTCATATCACACAGCTTTG
GGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTGTTTGCTGCTGCTCCTATGCTTTCTGGGGT
TGCTCCTAGTTCTGTGACTGAGGTGTTGGCTGATAGACAGTTACATGCTAGTATGCTGAAGGAGAAGTTG
GCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCATAGATCTGACAGAGAGTTTCAGATTGGA
GATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCGGTGGTTAGTCGACCCTGTCCTAAATTGG
CCTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGGGATGGTTGCTTATAAGTTGGATTTACCA
GTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGTTGAAGGCTTTTACTCCCAATTATTCTCCGGTG
TTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGATTATGCCATTGGCTGTGGTGGAAAGGAG
GATGGTGAAGAAGGGTAATCATCCTGTGGTGCAAGTAAGGGTGACTTGGTCCAATATGCCAGCTGAGGT
TACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTTCCTACAGCTCTTGATTGGGGACAATCAAGC
TCTTCCGGGGGGGGGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCT
GTTTTGCGTGGAGTTGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGT
AATAGTGGGCCGGCCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAA
ATAGAAACAGAATCCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGAT
TTGGCGCGGCTCACGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCACTCCAGTGATGATGCA
TGGT

CML333

>chr1:165200240-165203563

CTTGCTGTCATGACTGCTATGGTCGCTGACGAGGAAGTAGAAGCTATTAGTGTATAGAGATTCTGTAATG
GTGTGATGGGCTTTCAAGCCAAGATTTGTGTATAAGTAATTGGGCTGTTAAGCCAGGGATCAGTGAGTG
GGCCGGCCCAATAGACCAGGCAATATATACTGTAACCTGTCATCGGAACGGGGAGAGAAATATACAGAA
CTGAATAGAAGTTTTCTTACCTTCGTGCTTCGAGTTGCGTCTGTGTTCCTGGCGCTTCGTCGCCGAATTGG
TGTGGCTCACGTCGCCGGTGACCTTCCGGTGGCCAGGGGCGTGACAAGTGGTACCAAAGCTTCCGATCC
TCGGCGTCTCGGGGAGCGGTTCGTGTCTTCCTCATCGGCGTCGTTGCTGCGGTGACCAGGGTCTTCAACA
CGGGTTCTCGTCTCCATCGACGACTCTCGGGCTCTCGGCATCGGCATCGCATCATCGACGCTCGTCGATCT
CCGGTAGACAGTCGACTTGGTGGTGTTGGCGGCGGAGACCGGAGATAACCTGTTCGACGAGTTGCCTG
GGTGTCGTTCACTCGGCTGGGTCGTCGACGCAGGCTTCGTTCTCCATCGACGATCTGCGGCGTTTCCTCG
CTACAACGTCGTTGGTGATGTCGTCCGCTCAGACAAGCATCCTGCTCGCTGCGATGCCGACATCGGAGCA
GAAGGGCGAGGAACGATGGGACAAAGTGATGCGGGTCCTGGATCGCCTCGACAGTAGGTTGGAGGCG
ATGGAGGTGAAGCAGAAGGGTCTGCACGATCATACGTGCAACGTGGAAGCAGCAGTGCGGAACATCCC
AGCTTTGCTTGGAGGTATGGCCCCTCCTGTGTTCGATGAAATGCCCCCCTACTGCAGTGAGGAGGTCATT
GCCCTTCCAAGGTGCTTTGGAGGTTCTGGTTATTCAGTGTTCGATGAAACATCTGCTGAGAAGGTGGTCT
GGGATGAAGAAGTGTTACTTCGTTCCGATTCACTTGATTTTTGGCAGCGTCTAGCTCATGGAATGGGGAA
CCAGATGCCCAACGATTGCAAGCAGGATGTTGTGCCTGATGCCCCATGGTTGTTTGTTGGAATGTCACAG
CAAGTAAGTACATACTCAATATCCAGTGAAATGGGTCTTACAGTGGCATGGGATGGGGAGATGCAAGCT
GAAACTCACACGCATGAAGGATTACTGCAGCAGCTTGCTCAAGGTGAAGAAGCCTTAAAAGAATCAGCT

AAGAAGGAGGGTGACGTTGAAGATGAGGCAACCACAGTGGTTATGAATGGAGGAAACAGGGATGAGT
TTATTCCTGCTAGTGTGGTCAGTGTGCTTGAGGATGACCCTAAGCAGGATGCTGAAAATGTTGATGCTAC
CAGTGCTGGCGGTATTCTTTCTCAAGAATTGGCTCCAGAATCAATCAATGAAACTAATGGTGCAGGCAAG
ATTGTAGGCGTCGCTTGGGTTGATGATCATGAAGAAGAGGTTGTGGGTAGCGACATACATGAGGCTATG
ACAGATGAAGTGGATGAGGAAGAAGGCTGTTTTGAAGGTTTTTCAGCTATAAGAATACATGATGAAGAC
ATGAACCTGAAATTTCTTCTCTGTGGAGTACCATGCACCCTGGATGCATGTTTATTGGGCACACTAAAAGA
TGGCTTGAATGCACTCTTGAATATTGAGATTTGTGCGAGTAAACTTCAGAACCGAGCAGTGCTCATCCAC
CCCTTCAAGCTGAAGATCTTTCACATGGAGTTGCTACAATGCGCAGCGAAATCACAACTTGCAGTTCTGC
ACATGTGGCGTATTTTGTTTCTGGCAGTGCCCAAACTTGTTTTGGTGATCAACTTCTAGGGAGCCGTATCA
AGAATGAAATTATTGAGAAGAGGCAGTTAGTTTGTGTTCTATTGAATAATGAGGATATCAAGCCATCTTC
ATTTGAGCCTTGGCCATCCATGGGTGTGGCTTGTGAAGCTTCTACTTTTGAAGTATGGATAACCATGCAA
AAATGGGCAGCACAGGTTTTAAAACATCTTGCACCAGAAATCTCCTACAGAAGCTTGGTTGCGCTTGGAA
TCGCCTGGATAAATGGTATACCAGTGTCCTCATTTGGTAGGCAGGATGCAGACCAGTTTATTTTCTTTTGT
TCTAACCAGTGCGGAGATCAAGCTATTCTAAATGGCTCATTTGCTCATGTATCCAACTGGGCAGCACCCTT
TACTAAGGATAATTTCATGCTTGTTCTATATGAAGCATGTACTGGGCAAGAGGTGAGATTGGAATGCAAG
AATGTGTTGCTCTTTGAGGGGAGGCAAAACAAAGGTAGTTTTCGACATGGGATATCTGAACTGACACAA
TGGTGGCTAAATGGTTTTAGCCCAGGAAACGGTGCTATGGCAGTTCAGTCTCAGCCATACACTCAACAAA
AATTCATGAACAAATTAATGGGCTTACAGTTCAGAATGGTGTATGATAGAGGCAAGGAGAAGATAGCGA
CTGATGCTTTTGTTCAGGATTTGTGGCCGAGGGTGACCATTTACATTCCATATGGGAATGTGTTCTCAGTG
CAGAAGGACAAGACTAGAAGACTTCCTCAATTTTGGGTGGGCAGGAGCTCAACATTCAGCACTAAATTG
ATTCGAAGCTTTCATGATTCAATGGTGTGGGAGTGTTCTGTAAGCGCAGCGGAAGCATGGTGGATCGGT
CTGCAAACTATGGTGGAGCAGTTCGGAGAGCAGTGGTCAGTTACTGAGCAAGAGAAAGAAGAGCAGCA
ATACTTGTTTGCTATCCAGGATCATCCTCGACTGTGGAAAATTGGGGGTTGTTTTCCCTATACAGCACGAT
ATGGGTATAACAATGTGGTATTGGTTGTACTGGTGCTCAACGGGATTGTGGATTTAGTGGGGGCTGCTT
GGATACAACTGAAGGCTATGTGCTTAAGGAGAAATTTCCACATGCTCTTGCTGGAGGGCAAGCAAGTTTT
GCAGCAGGGGGAGATGTCATGACTGCTATGGTCGCTGACGAGGAAGTAGAAGCTATTAGTGTATAGAG
ATTCTGTAATGGTGTGATGGGCTTTCAAGCCAAGATTTGTGTATAAGTAATTGGGCTGTTAAGCCAGGGA
TCAGTGAGTGGGCCGGCCCAATAGACCAGGCAATATATACTGTAACCTGTCATCGGAACGGGGAGAGAA
ATATACAGAACTGAATAGAAGTTTTCTTACCTTCGTGCTTCGAGTTGCGTCTGTGTTCCTGGCGCTTCGTC
GCCGAATTGGTGTGGCTCACGTCGCCGGTGACCTTCCGGTGGCCAGGGGCGTGACACTTGC

HP301

>chr5:123933009-123936923

TGGTTCGATTGCTCCATGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTTGCGTGGAGT
TGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATAGTGGGCCGG
CCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAGAAACAGAAT
CCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGGCGCGGCTCA
CGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCATCTGGTATCGGATTCGGGGTCGGCTGCGT
CTTCATCCTCTTCCACGTGCTGTTCCTCGTCTTTGGCGGTTCGTGGGCGACTTCTTCTTCTGCTGGGTCCAT
CCAGTTCATCACTACGGCGCCCGCAACCAAACCATCGGTACAGACTCGTTTGGTGTTACAGGCTATGGCG
GATGGGGAGCGAGTTAGTCAGGAGAATTGGAGTCGTGTTATTGAGACGTTGGAGGGAGTGAGTAGCAG
GTTGAAGGATGTGGAGAGGGTGCAGCAACAAATCATGATGCGGACTGAATTAGCGGCATCAGTTGCTG
AACAGGCTGCGAAGGAGAGAAATGATTTGGGTCGTAAATTGGAGGAGACTGGAAGGGGAATGATATTT
GGGCTTTAGTGGTGTTGGAAGAGGTGTCTGCAGATCAGGGAGTACAGCAGCAGGGGGATCTTCAGCAG

TTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTACCTCCTGCTAGGGTGTATGATCATCA
TATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTATAGGTACTCTCCTCTTCACAAGGATG
AAATTGAGAGACAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATTGTTCCTAGTGTAAGCCCGTTTGCTTCT
CCGGTATTGTTGGTTAAGAAGAAAGATAGCACTTGGAGGTTTTGTGTGGACTACAGGAAATTAAATTCTT
TGACCATTAAGAACAGATTTCCTATGCCGCTTGTAGAAGAGATTCTGGATGAACTGGCTGGTACTAAATA
TTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAATGGGGGAAGGTGATGAGCATAAAACA
GCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCCTTTTGGGTTAACCAATGCTCCTGCAAC
CTTTCAATGTGCTATGAATTCTGTTTTGGCTCCTTTTCTTCGGAAATTTGTCATGGTTTTCATTGATGACATT
TTGGTCTATAGTGCTACATGGGAGGAACACTTGGCTCATATTAAGATGGTGCTGCAGACTTTGAGAGAG
CACCAGTTCTATCTGAAGAGGAGTAAATGTGTGTTTGGGAGGAATGAGCTCATTTATTTGGGACATATTA
TCTCTCACACAGGAGTGGCCACTGATCCTGCTAAGACAACAGCCATGGAGAAATGGCCTGTTCCTCAGTC
AGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTATAGGAGATTTGTGAGACATTATGGGAT
TATTGCAAAACCTCTTACTAATCTGTTGAAGCATAAGGCTTTTCAGTGGAGTCCTGAGGCTGATCAGGCTT
TTCAAGCGCTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCCCTACCCAGGTTTGATGAGTTATTTGTT
GTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGATGCAATGTGATAGACCAGTGGCATAT
CTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTATCCATTTATGAAAAAGAGTTTTTGGCTCTAATCA
TTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATTTGTGATTAAAACTGATCATCAGGCTT
TGAGTTTTCTGGATGATCAGCTTTTGCAATCTGATATGCAGAGGAAGGCTATGGCTAAGTTGATGGGGCT
TCATTTCAAAATTGTTTACAAAAAAGGCAAGGACAATGTGGCTGCCGATGCTCTATCCAGGATTGGTCAT
GTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGGGTGCAAGAACTCCTGAATGCTTATGTG
ACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGGCCAGTCCTAATGAGCAAGGTTTTTCTT
TACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGCTAACAATTCTGCCCTTCAAACTAAGTT
GATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGTGGTGAGGCCACTTACCAAAGGATCAA
GAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAGTTTGTGCAGCAATGTGAGGTGTGTCA
GCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAACCTTTACCAGTGCCTCAAGGGGCTTGG
CAGGATCTTACATTGGATTTCATTGAAGGTCTACCTAGATCTGAGGGATATGATACTATTTTGGTCGTGGT
GGATAGATATTCCAAATATGCTCATTTTTTTCCCCTCAAGCATCCTTTCACTGCTGTTGGAGTGGCACAGG
TGTTTCTTGACAATATTGTCAAGTTGCATGGGACACCAAAGTCTTTGGTTTCCGATAGAGACGAAGTGTTC
ACTAGTTTGTTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCAAATTGGCTCTCACTACTGCATATCACCCT
CAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGAAATGTATCTTCGCTGTTCTATTCATGAGG
CACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAATTTTGGTACAATACTTCATATCACACAGCTTTG
GGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTGTTTGCTGCTGCTCCTATGCTTTCTGGGGT
TGCTCCTAGTTCTGTGACTGAGGTGTTGGCTGATAGACAGTTACATGCTAGTATGCTGAAGGAGAAGTTG
GCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCATAGATCTGACAGAGAGTTTCAGATTGGA
GATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCGGTGGTTAGTCGACCCTGTCCTAAATTGG
CCTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGGGATGGTTGCTTATAAGTTGGATTTACCA
GTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGTTGAAGGCTTTTACTCCCAATTATTCTCCGGTG
TTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGATTATGCCATTGGCTGTGGTGGAAAGGAG
GATGGTGAAGAAGGGTAATCATCCTGTGGTGCAAGTAAGGGTGACTTGGTCCAATATGCCAGCTGAGGT
TACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTTCCTACAGCTCTTGATTGGGGACAATCAAGC
TCTTCCGGGGGGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTT
GCGTGGAGTTGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATA
GTGGGCCGGCCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAG

AAACAGAATCCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGG
CGCGGCTCACGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCACTCCAGTGATGATGCATGGT

La453

>Chr5:124841462-124845381

TGGTTCGATTGCTCCATGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTTGCGTGGAGT
TGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATAGTGGGCCGG
CCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAGAAACAGAAT
CCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGGCGCGGCTCA
CGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCATCTGGTATCGGATTCGGGGTCGGCTGCGT
CTTCATCCTCTTCCACGTGCTGTTCCTCGTCTTTGGCGGTTCGTGGGCGACTTCTTCTTCTGCTGGGTCCAT
CCAGTTCATCACTACGGCGCCCGCAACCAAACCATCGGTACAGACTCGTTTGGTGTTACAGGCTATGGCG
GATGGGGAGCGAGTTAGTCAGGAGAATTGGAGTCGTGTTATTGAGACGTTGGAGGGAGTGAGTAGCAG
GTTGAAGGATGTGGAGAGGGTGCAGCAACAAATCATGATGCGGACTGAATTAGCGGCATCAGTTGCTG
AACAGGCTGCGAAGGAGAGAAATGATTTGGGTCGTAAATTGGAGGAGACTGGAAGGGGAATGATATTT
GGGCTTTAGTGGTGTTGGAAGAGGTGTCTGCAGATCAGGGAGTACAGCAGCAGGGGGATCTTCAGCAG
TTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTACCTCCTGCTAGGGTGTATGATCATCA
TATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTATAGGTACTCTCCTCTTCACAAGGATG
AAATTGAGAGGCAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATTGTTCCTAGTGTAAGCCCGTTTGCTTCT
CCGGTATTGTTGGTTAAGAAGAAAGATAGCACTTGGAGGTTTTGTGTGGACTACAGGAAATTAAATTCTT
TGACCATTAAGAACAGATTTCCTATGCCGCTTGTAGAAGAGATTCTGGATGAACTGGCTGGTACTAAATA
TTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAATGGGGGAAGGTGATGAGCATAAAACA
GCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCCTTTTGGGTTAACCAATGCTCCTGCAAC
CTTTCAATGTGCTATGAATTCTGTTTTGGCTCCTTTTCTTCGGAAATTTGTCATGGTTTTCATTGATGACATT
TTGGTCTATAGTGCTACATGGGAGGAACACTTGGCTCATATTAAGATGGTGCTGCAGACTTTGAGAGAG
CACCAGTTCTATCTGAAGAGGAGTAAATGTGTGTTTGGGAGGAATGAGCTCATTTATTTGGGACATATTA
TCTCTCACACAGGAGTGGCCACTGATCCTGCTAAGACAACAGCCATGGAGAAATGGCCTGTTCCTCAGTC
AGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTATAGGAGATTTGTGAGACATTATGGGAT
TATTGCAAAACCTCTTACTAATCTGTTGAAGCATAAGGCTTTTCAGTGGAGTCCTGAGGCTGATCAGGCTT
TTCAAGCGCTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCCCTACCCAGGTTTGATGAGTTATTTGTT
GTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGATGCAATGTGATAGACCAGTGGCATAT
CTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTATCCATTTATGAAAAAGAGTTTTTGGCTCTAATCA
TTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATTTGTGATTAAAACTGATCATCAGGCTT
TGAGTTTTCTGGATGATCAGCTTTTGCAATCTGATATGCAGAGGAAGGCTATGGCTAAGTTGATGGGGCT
TCATTTCAAAATTGTTTACAAAAAAGGCAAGGACAATGTGGCTGCCGATGCTCTATCCAGGATTGGTCAT
GTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGGGTGCAAGAACTCCTGAATGCTTATGTG
ACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGGCCAGTCCTAATGAGCAAGGTTTTTCTT
TACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGCTAACAATTCTGCCCTTCAAACTAAGTT
GATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGTGGTGAGGCCACTTACCAAAGGATCAA
GAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAGTTTGTGCAGCAATGTGAGGTGTGTCA
GCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAACCTTTACCAGTGCCTCAAGGGGCTTGG
CAGGATCTTACATTGGATTTCATTGAAGGTCTACCTAGATCTGAGGGATATGATACTATTTTGGTCGTGGT
GGATAGATATTCCAAATATGCTCATTTTTTTCCCCTCAAGCATCCTTTCACTGCTGTTGGAGTGGCACAGG
TGTTTCTTGACAATATTGTCAAGTTGCATGGGACACCAAAGTCTTTGGTTTCCGATAGAGACGAAGTGTTC

ACTAGTTTGTTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCAAATTGGCTCTCACTACTGCATATCACCCT
CAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGAAATGTATCTTCGCTGTTCTATTCATGAGG
CACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAATTTTGGTACAATACTTCATATCACACAGCTTTG
GGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTGTTTGCTGCTGCTCCTATGCTTTCTGGGGT
TGCTCCTAGTTCTGTGACTGAGGTGTTGGCTGATAGACAGTTACATGCTAGTATGCTGAAGGAGAAGTTG
GCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCATAGATCTGACAGAGAGTTTCAGATTGGA
GATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCGGTGGTTAGTCGACCCTGTCCTAAATTGG
CCTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGGGATGGTTGCTTATAAGTTGGATTTACCA
GTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGTTGAAGGCTTTTACTCCCAATTATTCTCCGGTG
TTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGATTATGCCATTGGCTGTGGTGGAAAGGAG
GATGGTGAAGAAGGGTAATCATCCTGTGGTGCAAGTAAGGGTGACTTGGTCCAATATGCCAGCTGAGGT
TACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTTCCTACAGCTCTTGATTGGGGACAATCAAGC
TCTTCCGGGGGGGGGGGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGC
TGTTTTGCGTGGAGTTGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTG
TAATAGTGGGCCGGCCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGA
AATAGAAACAGAATCCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGA
TTTGGCGCGGCTCACGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCACTCCAGTGATGATGC
ATGGT

>Chr9:20973591-20978517

CCTCCTGTTACCAGCCAGGCGTCAGTATCTGATCACCCACAGAAGCAGCCCCAGCGTCGCAAGCCTGGAA
CGCGTGAGCGTCGTGCCAATGTCCGCGTCACAGGACCGGAGTGGGTGTAGTTGTGTGCGTGGTGTTCCT
GTCCCTGTTTGATGGACCAGCTGCGTCTGGTGGAGGGTTTCTTTATGTATCGCTAACTCTGAACTGGGAA
GGCAGGAAACAAACAGAGAGAAAATATCTATTCCGGCCAGTTCCTATCTCTCTCTCTGTCGTCCTCCTCCT
CCAACCTTTTCCCCATTTCCTTCTGTTAGTATAGATCAGAGCTGCTAACAAATTGGTATCAGAATCCTCCGT
TCCTCGGCCGTCGCGTCCATGGCGTCTCCTGTCGACGCGCAGCTCACGCAGGCGCTGGAACAACAGACCC
GCGCCATTCTCGCTGAGCTCGACAAACGGTTCACAGCGATCGACTCCAAGTGGGAACTCCGCGTCGGCG
CGCTGGAGTCCCAGGCGGCGGACTTCGCTCGCCACTCCGACGAGTTCGCCTCTGCCATCCGCGCGGACG
TCGAAGCTCACCTTTCGGCGGCCGATGCGGATCTTACCGATCGCGTCCGCTACTGGGAGGCCACTACCGG
AGCGCGCGTCACCTCCCTCGAGTCCGCCCAGCACACGTTCGACGCTTGGCGCCCCAGGATGGAATCCTCC
GTCGACGCCCTACAATCTTCCATGGATTCCGTTCGCGCCGAGCTATCCAAGGTGGAGATCCGGTGGTCCC
AAGGGGCGCGCGTCGAGGGCTTCTCCAAGCCCGGCCTCCTCGGCTCGCCCAGATCGGCGCCGGTGTGTT
CTTCTGCCACCGGCGGTCAAGCTGACGGCTCACGTTTTGGGCCCCGCTTCGCTTCATCTCACCGGGAATG
TGGATTTGGGCCCCCTGTGGCCGATCTCCACTACCCGGTCAAGGGTACGTGGTCACCTCACGCCCCTGCT
TCGTTTGAGTTTCCAGTTCATCATAATCTTAATTCTGGGGAGCTAGATCGTCGGGGGGAACTGAGTAGAG
GATTGGGTGGTCTTCCTAAGGTCAATTTCCCTGTGTTCGATGGCAGTAATCCCAAACTTTGGCAAAAACA
GTGCGAAGATTATTTTGACATGTACGCTACCGAGGAGTTCGTTTGGATTAAAGCCGCAACTATGCATTTC
CATGGCGTAGCTGCGCGCTGGCTCCAATCCATCGAACCTCGCCTGCCCTCGCTCTCCTGGCGACAGTTCT
GTCAGTTCGTCCATGACAGATTTGGTCGTCAACAACATGAGCTAGTTATCCAAAAATTATTCCACATCAAA
CAGACCTCCACTGTCCAGGATTACATTGATCGCTTTTGCGAGTTAGTGGATTTGCTTGTCACCTATGAACA
CACAACAGACCCATTGTATTACACTATGAGATTTATTGATGGCTTGCGTGCTGAAATCAAGTCTGTCATTC
TTGTACAACGTCCTGGCGACCTCGATACTGCCTGCTCTTTGGCCTTGTTGCAGGAAGCTGCAGAGTCTTCT
CGTCGGCCGGAATTGGTTCGGGGTGACCATTCTTCGTTTTCTCGGCCGCTGCTAAGAGCCGCTCCCACGT

CGTCCTACCGTGGGGAGCAGCCTCGTGGTGGCGGGGGCGATTATACCAAGACCAACACGACGGGGCCT
ACAACTGTTGATTCTAAGGTGGCCTCTCTTCGTGCGTACCGGCGTGCTAAGGGTCTCTGTCAATACTGCG
CAGAGAAATGGAACAAGGGTCATAAATGTGCAGATACAGTGCAGTTGCATGCTGTCCAAGAATTGTGGG
AGATGCTGTCAACTGATCAGCCTGAATCTGAAGGAGAGTTTGAGGATTCTGCTGAACAATTTATGATATT
ATTGTCTACAGAAGCTGCCTCGGTTCACCGCCCATCAAAATCATTCAGAATTCGAGGTGTCCTGCAAGGG
GTTGACCTGCTCATGTTGATCGACTCAGGTAGTTCTCATAGCTTCCTGAATTCTGCCCATACTTCTAAGTTG
TCTGGCATCACTTCTATGGATACACCATTGGCTGTACGAGTAGCAAATGGTACTGTGTTGCAGTGTTCTGT
GGAACTCCCTCAGGCTTCTTGGTCCGTGCAGGATCTTAGCTTCTGTACTACATTCAAGTTGATCCATCTAC
CCTTCTATGATGCAATATTGGGTATGGATTGGCTGGAACAGCTGAGCCCAATGACAGTGGACTGGAAAC
ACAAATGGTTGTCCATCTCATATGCAGGCAGCACAGTAATTTTACAAGGTTGTAATCCTTTGGTTCCGGTT
GGTACAGTGTTGGAAATTCAACGTTTGCTTCCTGTTTCGACTGATCCTGCTCAGTTGCACTCGGTCGAAGA
CTTGCAGTTACCAGAGGCAGTGATGCGACTGCTTCGCAACTTTCAGGATGTGTTCCAGCCACCATCTGCC
CTGCCCCCGACTCGATCGTGTGATCATGCTATCCCTTTGATTGAAGGAGCGCGACCAGTCAACATCCGAC
CCTATCGATTTTCACCAGCAATGAAAGATGAGATGGAGTCTCAGGTGGCTGAAATGTTGCAGAGTGGTTT
GATTCAGCACAGTGATAGTGCATTCTCATCCCCAGTGTTGCTGGTCAAAAAGAAGGATAATTCATGGCGT
TTTTGCGTGGATTACAGACATTTAAACGCCCTTACAGTTAAATCCAAGTACCCTGTTCCCATCATCGATGA
GTTACTTGACGAGCTGTTTGGTGCCTCCTGGTTCTCCATATTAGATCTAAGAGCGGGCTTCCACCAGATCT
TACTCAAAGCAGGAGAGTCACACAAGACTGCCTTCCAGACACACACAGGCCACTATGAGTTTCGAGTTAT
GGCTTTTGGTCTCACTGGAGCTCCTGGTACATTCCAGAAGGCTATGAACCACACCCTAGCTCCCCTGCTTC
GGAAATGTGCTCTTGTGTTTTTTGATGACATTCTGGTTTACAGTGCCTCCTTGGAAGAGCATATTCATCAC
TTACAACTGGTTTTCGAGTTGTTAGCCCAGGATCAATGGAAAGTTAAGCTCTCCAAATGTTCTTTTGCACA
ACAGCAAGTCTCCTATTTGGGCCATGTCATCAATGCTCAAGGTGTAGCTACAGATCCAGTTAAAATTCAA
GCTATTGCCGAGTGGCCAGTGCCAAACTGTCTCAAGGATTTACGCAGTTTTTTGGGCCTTGCCGGCTACT
ACAGGAAGTTTATTAGGCATTTTGGGGTCATTTGCCAGCCCTTGACCGCTTTACTGAAGAAGGGCAGTCT
GTTTGTTTGGACAAATGAGCATGGTATTGCATTTGACACTTTAAAAAAGGCTTTGGTGACAGCTCCTGTTC
TTGCATTACCTGATTTCACTGCTGTTTTTCTCGTGGAAACTGATGCTAGTGCTGTCGGTGTGGGTGCAGTC
CTCATGCAAAAAGGTCATCCCCTCGCCTTCCTTAGTAAGGCCCTGGGTCCCAAGTCTAGGGGATTGTCAA
CATACGAAAAAGAGTACTTAGCAGTGATCTTGGCTGTTCAACAGTGGCGCCCCTACCTGCAACACCAGGA
GTTTGTTATCCTCACAGACCATAAAAGCCTCACACAGCTTAATGAGCAGCGCCTCCATACTCCTTGGCAAC
ACAAGGTATTTACCAAGCTCCTTGGTTTGCAGTACAGGGTCCAATACCGTCCGGGCACTGAGAATCGTGT
TGCTGATGCGTTGTCGCGGTGTGCCACCGCTGATCTGCACGCTCTGTCATCAGTCGTGCCTCAGTGGTTA
CTGGACGTGCAGGCCAGTTATACACAGGATGCTGCAGCTCAATCCATTCTGGCCAAACTTGCAGTCGACC
CAGCGGCTCTTCCAAACTATGTACTGAAGGATGGACTGATTCGGTATAAAGGTCGTATATGGTTGGGTGC
CGATGAACAACTACACCGCCGAGTTGTTTCTGCGTTACATTGCTCAGCCGTCGGTGGCCACTCGGGCTTC
CCTGTCACTTATAGCCGAGTCAAGCATCTGTTCGCTTGGACCAAGCTCAAGTCATTTGTTCAACGCTTTGT
CCAGCAGTGTCAGATCTGTTTACAAGCTAAACCGGACAGATCAACTCAGGATGCAGCGTCAGGCCGACA
AAAACAGGTCTGAGCGCTCGTTTGCCATCGGTGATTGGGTGTACATGAAACTCCAGCCCTATGTGCAGTC
ATCTGTCTCGCTGCGAGCTCATCATAAGCTGAGTTTTAAGTATTTTGGCCCATACAAAGTCACAGCTCGGG
TTGGTAAGGTTGCATATCGTTTGGACCTGCCTGTCTCCAGTAAGATCCACCCGGTGGTCCACGTATCACA
ACTGAAACTGGCAGCAGGTTACAAAGGCCCTGCTCCAGCAACTCTGCCTCTGGATTTACCTGAATTCAGT
GTTCCTGTTCAAATCCTTCAGACGCGAGGACTTACCAAAGGCAACCGTCTCGTACAGCAAGGCCTTGTGC
AGTGGTCCGGCTTACCTTCCGACCTGGCCACCTGGGAAGATCTTGAAGCTTTGCGTCAATGTTTTCCTGGT
GCTCCTGCTTGGGGTCAAGCAGGATTTCAAGGGAGCGGGAATGTTACCAGCCAGGCGTCAGTATCTGAT
CACCCACAGAAGCAGCCCCAGCGTCGCAAGCCTGGAACGCGTGAGCGTCGTGCCAATGTCCGCGTCACA

GGACCGGAGTGGGTGTAGTTGTGTGCGTGGTGTTCCTGTCCCTGTTTGATGGACCAGCTGCGTCTGGTG
GAGGGTTTCTTTATGTATCGCTAACTCTGAACTGGGAAGGCAGGAAACAAACAGAGAGAAAATATCTAT
TCCGGCCAGTTCCTATCTCTCTCTCTGTCGTCCTCCTCCTCCAACCTTTTCCCCATTTCCTTCTGTTAGTATAG
ATCAGAGCTGCTAACACCTCC

Il14H

>chr9:20170887-20174667

AAAATAAGAAGGTGAGCCTGTCATGTGGGGTCTCTCCTATACGCGCAGGAGGCGCACCAGGGATGTTAG
GCGTGCGGCCGAGAGGGCTGCAGCGCGCCCACCGGCCCTGCAACAGGAAGGCCGGAATTCTAGTGGAT
AAGATAGGCAATAGAATTAGTTGAGGATCCACAAAATTAGGAGGGAATTCTAGTGGATAAGATAGGCAC
TAGAATTAGTTGAGGATCCACAAAATTAGGAGGTTTGTTGCATGCGGCAGGCTATTAAAGCCTTGTACTG
GGCGCTGATCAGGATATTAAGAAATAACACAGCTATCTTCTTCCTCCTCCCTACCAATTCCCAAGCCTTCG
GAAGAGGGGTAAAACCTCACCGGAGAAGACGGACCGAGGCTACAACCTGCGTAGCCGAGGGCCTCCTA
CTCTAGACACTAAGGCCCTTGACAACCTGGTATCACGGTCACGTCGATCCAACTCTTCCAGTTCATCCGCT
GCATCCACTGCTGCTTTTCCATCTGCCGCACCCACCACACTGCGGCCAGCGACTCCCTCCTCGACGGGCTG
CTCGGTACGTTTACCGACGTCTTCGCCGCTCCGACGGGGCTACCACCGGTGCGTGGTCGCGACCACGCCA
TCGTCCTCAAGCCGGGTGCAGCGCCGGTGGCTGTCCGTCCCTACCGATACCCGGCGGCACACAAGGACG
AGCTGGAGCGGCAGTGCGCCGAGATGATGGGTCAAGGCATCGTTCGCCGGAGCGACTCGGCGTTCTCG
TCGCCAGTCATCCTCATCAAAAAGCAAGACGGCTCCTGGAGGTTCTGCGTCGACTACAGGGCGCTGAAC
GCGCTCACCGTCAAGGACGCCTTTCCAATTCCCGTCGTCGACGAGCTGCTGGACGAGCTCAACGGCGCCC
GTTTCTTCACTAAATTGGACCTGCGATCAGGGTACCATCAGGTGCGCATGAAGACGGCAGATATTCACAA
GACGGCCTTTCGAACCCACGACGGCCTCTACGAGTTCCTCGTCATGCCGTTCGGCCTTTGCAACGCGCCG
GCGACGTTCCAGGCCCTCATGAACGACGTCCTCCGGCCCTTCCTCCGTCGGTTTGTTCTAGTATTTTTGA
CGACATCCTGATATACAGTTCTTCATGGGCCGACCACCTGCGCCACATTCGTGCGGTGCTGACGGTGCTG
CAGCAGCACCAACTCTTCGTCAAGAAGTCCAAATGCGCATTTGCCGTGCCATCGGTGGCATACTTGGGCC
ATGTCATCTCGGCGTCCGGCGTGGCCATGGATCCAGCCAAGGTGGAGGCCGTCCAGGGCTGGCCACAGC
CCCGTTCGGCGCGGGCGGTCAGGGGATTCTTGGGCCTTGCGGGGTACTACCGCAAGTTCGTCCACAACT
ACGGCGTCATCGCGGCTCCCCTGACGGCGCTCCTCAAGAAAGAAGGCTTCTCTTGGACGGAAGAGGCAA
CTGCAGCGTTTGCAGCGCTAAAGGCAGCGGTTACATCGGCGCCGGTCCTCGCGCTGCCGGACTTCACCA
AGCCGTTCATCGTCGAGTGCGACGCGTCGACCCATGGCTTTGGGGCGGTGCTGATCCAGGACAGCCACC
CTATCGCTTTCTTCAGCCGGCCCGTGGCGCCCAGGCACCGCTCACTCGCTGCATACGAGCGGGAGCTCAT
CGGACTTGTCCACGCTGTCCGTCACTGGCGACCGTATCTATGGGGGCGTCGGTTCATCGTCCGCACCGAC
CACTATAGCTTGAAGTTCCTCCTCGACCAGCGGCTAGCTACCATCCCTCAACATCATTGGGTGGGAAAGC
TGCTGGGGTTTGATTTCGGTGTCGAGTACAAGCCGGGCGCCCAGAACGCGGTGGCGGACGCCCTGTCAC
GTCGAGACACGCCGGATGGAGCGGTGCTCGCCATATCGGCGCCTCGCTTCGACTACATTGCCAGGCTGC
GGCTGGAACAGGCCGGCCATCCGGCTGTGGTCGCCCTCAAGGACGAGTTGTCAGCAGGATCGCGCGGC
GCACCCTGGGGCATTGTCGACGGGCTGCTCACCTACGACAGCCGTATCTTCGTTCCGCCGGCGTCACCGC
TCCTACAGGAGATCCTGGCTGCTGTCCATGAGGACGGCCACGAGGGCGTGCAGCGCACCCTCCATCGCC
TTCGCCGAGACTTCCACTTTCCCGACATGCGCCGCGTGGTGCAGGAGTTCGTCCGCGCATGCTCGACTTG
CCAGCGCAACAAGTCGGAGCATCTCCTGCCGGCTGGGTTGCTGCTGCCCCTGCCGGTCCCGTCAGGCGT
GTGGACGGACATCGGCCTCGACTTCATGGAAGCTCTACCGCGTGTGCGCGGCAAGTCGGTCATATTGAC
GGTGGTGGATCGCTTCAGCAAGTACGCGCATTTCATTCCACTGGCTCATCCGTATTCAGCTGAGACGGTC
GCACAGGTGTTCTTCTCCGACATTGTACGCCTCCATGGCGTGCCTCAGTCGATCGTATCCGATCGTGATGC
TGTTTTCACCTCCACGTTCTGGCAAGAACTGATGAGGCTCATGGGCGCCAAGCTTCACATGACGTCGGCA

TTTCACCCACAGTCTGATGGTCAGTCGGAGGCGGCCAACAAGGTCATCGCCATGTACCTCCGCTGCTTCA
CAGGAGACCGGCCGCGCGAGTGGCTCAGGTGGCTGCCCTGGGCGGAGTACACGTACAACACGGCGTAC
CAGTCTTCGCTCCGCGACACACCGTTCCATGTGGTGTACGGCCAGCATCCCCCCTCCATCCGCTCCTATGA
GGTGGGTGACACCCGCGTGGCCGCTGTGGCCAAGACGATGGCCGAGCGGGAGGCGTTCCTGGAGGAT
GTGCGCCTCCGCCTCGAGCAGGCTCAGGCTGTGCAGAAGAAGTTTTACGACAACAAGCACAGGGCGGTG
TCCTATACCGTTGGGGACTGGGTTCTGCTGCGCCTTCGCAACCGGCCGGTGGCCTCCCTATCCCTGGCAC
TGAAAGGGAAGCTGCAGCCTCGCTTCTTCGGGCCCTACCGCGTCACCGAGATCATCAACGACGCTGCAGT
ACGCCTCGCCCTCCCGCCGCAGGCCAAGATTCATGATGTGTTTCACGTGGGACTCTTGAAGAAATGGGTG
GGCGCGCCGCCTGCCTCGACGCCGGAGCTGCCCGAGTTGCTGCATGGGGCGGTGGTGCCCGAACCCGA
ACGAGTGGTGCGTTCCCGCATAGCTCGAGGCATCAAGGAGGTGTTGGTGCGCTGGAAGGGACAGCCGA
CTTCTGCTGCTACTTGGGAAGACTGGGACGACTTTATCGCCAGGCACCCGTCGTTTCAGCTCGAGGACGA
GCTGCGAGTCGAGGGGGGGAGAGATGTCATGTGGGGTCTCTCCTATACGCGCAGGAGGCGCACCAGGG
ATGTTAGGCGTGCGGCCGAGAGGGCTGCAGCGCGCCCACCGGCCCTGCAACAGGAAGGCCGGAATTCT
AGTGGATAAGATAGGCAATAGAATTAGTTGAGGATCCACAAAATTAGGAGGGAATTCTAGTGGATAAG
ATAGGCACTAGAATTAGTTGAGGATCCACAAAATTAGGAGGTTTGTTGCATGCGGCAGGCTATTAAAGC
CTTGTACTGGGCGCTGATCAGGATATTAAGAAATAACACAGCTATCTTCTTCCTCCTCCCTACCAATTCCCA
AGCCTTCGGAAGAGGGGTAAAACCTCACCGGAGAAGACGGACCGAGGCTACAACCTGCGTAGCCGAGG
GCCTCCTACTCTAGACACTAAGGCCCTTGACAGAGCCCCAAAA

>chr5:122944015-122947932

TGGTTCGATTGCTCCATGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTTGCGTGGAGT
TGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATAGTGGGCCGG
CCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAGAAACAGAAT
CCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGGCGCGGCTCA
CGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCATCTGGTATCGGATTCGGGGTCGGCTGCGT
CTTCATCCTCTTCCACGTGCTGTTCCTCGTCTTTGGCGGTTCGTGGGCGACTTCTTCTTCTGCTGGGTCCAT
CCAGTTCATCACTACGGCGCCCGCAACCAAACCATCGGTACAGACTCGTTTGGTGTTACAGGCTATGGCG
GATGGGGAGCGAGTTAGTCAGGAGAATTGGAGTCGTGTTATTGAGACGTTGGAGGGAGTGAGTAGCAG
GTTGAAGGATGTGGAGAGGGTGCAGCAACAAATCATGATGCGGACTGAATTAGCGGCATCAGTTGCTG
AACAGGCTGCGAAGGAGAGAAATGATTTGGGTCGTAAATTGGAGGAGACTGGAAGGGGAATGATATTT
GGGCTTTAGTGGTGTTGGAAGAGGTGTCTGCAGATCAGGGAGTACAGCAGCAGGGGGATCTTCAGCAG
TTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTACCTCCTGCTAGGGTGTATGATCATCA
TATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTATAGGTACTCTCCTCTTCACAAGGATG
AAATTGAGAGGCAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATTGTTCCTAGTGTAAGCCCGTTTGCTTCT
CCGGTATTGTTGGTTAAGAAGAAAGATAGCACTTGGAGGTTTTGTGTGGACTACAGGAAATTAAATTCTT
TGACCATTAAGAACAGATTTCCTATGCCGCTTGTAGAAGAGATTCTGGATGAACTGGCTGGTACTAAATA
TTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAATGGGGGAAGGTGATGAGCATAAAACA
GCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCCTTTTGGGTTAACCAATGCTCCTGCAAC
CTTTCAATGTGCTATGAATTCTGTTTTGGCTCCTTTTCTTCGGAAATTTGTCATGGTTTTCATTGATGACATT
TTGGTCTATAGTGCTACATGGGAGGAACACTTGGCTCATATTAAGATGGTGCTGCAGACTTTGAGAGAG
CACCAGTTCTATCTGAAGAGGAGTAAATGTGTGTTTGGGAGGAATGAGCTCATTTATTTGGGACATATTA
TCTCTCACACAGGAGTGGCCACTGATCCTGCTAAGACAACAGCCATGGAGAAATGGCCTGTTCCTCAGTC
AGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTATAGGAGATTTGTGAGACATTATGGGAT

TATTGCAAAACCTCTTACTAATCTGTTGAAGCATAAGGCTTTTCAGTGGAGTCCTGAGGCTGATCAGGCTT
TTCAAGCGCTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCCCTACCCAGGTTTGATGAGTTATTTGTT
GTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGATGCAATGTGATAGACCAGTGGCATAT
CTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTATCCATTTATGAAAAAGAGTTTTTGGCTCTAATCA
TTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATTTGTGATTAAAACTGATCATCAGGCTT
TGAGTTTTCTGGATGATCAGCTTTTGCAATCTGATATGCAGAGGAAGGCTATGGCTAAGTTGATGGGGCT
TCATTTCAAAATTGTTTACAAAAAAGGCAAGGACAATGTGGCTGCCGATGCTCTATCCAGGATTGGTCAT
GTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGGGTGCAAGAACTCCTGAATGCTTATGTG
ACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGGCCAGTCCTAATGAGCAAGGTTTTTCTT
TACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGCTAACAATTCTGCCCTTCAAACTAAGTT
GATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGTGGTGAGGCCACTTACCAAAGGATCAA
GAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAGTTTGTGCAGCAATGTGAGGTGTGTCA
GCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAACCTTTACCAGTGCCTCAAGGGGCTTGG
CAGGATCTTACATTGGATTTCATTGAAGGTCTACCTAGATCTGAGGGATATGATACTATTTTGGTCGTGGT
GGATAGATATTCCAAATATGCTCATTTTTTTCCCCTCAAGCATCCTTTCACTGCTGTTGGAGTGGCACAGG
TGTTTCTTGACAATATTGTCAAGTTGCATGGGACACCAAAGTCTTTGGTTTCCGATAGAGACGAAGTGTTC
ACTAGTTTGTTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCAAATTGGCTCTCACTACTGCATATCACCCT
CAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGAAATGTATCTTCGCTGTTCTATTCATGAGG
CACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAATTTTGGTACAATACTTCATATCACACAGCTTTG
GGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTGTTTGCTGCTGCTCCTATGCTTTCTGGGGT
TGCTCCTAGTTCTGTGACTGAGGTGTTGGCTGATAGACAGTTACATGCTAGTATGCTGAAGGAGAAGTTG
GCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCATAGATCTGACAGAGAGTTTCAGATTGGA
GATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCGGTGGTTAGTCGACCCTGTCCTAAATTGG
CCTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGGGATGGTTGCTTATAAGTTGGATTTACCA
GTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGTTGAAGGCTTTTACTCCCAATTATTCTCCGGTG
TTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGATTATGCCATTGGCTGTGGTGGAAAGGAG
GATGGTGAAGAAGGGTAATCATCCTGTGGTGCAAGTAAGGGTGACTTGGTCCAATATGCCAGCTGAGGT
TACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTTCCTACAGCTCTTGATTGGGGACAATCAAGC
TCTTCCGGGGGGGGGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTG
TTTTGCGTGGAGTTGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTA
ATAGTGGGCCGGCCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAA
TAGAAACAGAATCCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATT
TGGCGCGGCTCACGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCACTCCAGTGATGATGCAT
GGT

K0326Y

>chr4:145075504-145080291

CAGAGTGTTAAAAGATAGAAGGATATATCCATCCTTGGGAGCTGCGATCTCCCTGATTTCCTGAACCAAT
CTATTCTTGGGAAGTTATTATCTCCCTGATTTCTTGGAGCTACACCTCCGTGATTCTCTAGAGTTGTTACCA
TGTAAATCCTGTGTACCTCGTTCCTATATATTGAAGGATGCCGTGACCCCTAGAGGGCAGGCTGTGCATT
GCGTCAACATGGTATAGGAGCCAACTATCGATCACAGATGTCCACAGACTCCTCTCTTTTTTCCGCCTCCT
CCCCTTCCACCATGACTACCAACGCTCCAGCATTCCAATCTGCCTGCGTTCCAATTCCTCTGCACCATGTGG
TAACCATCCGTCTTACCAAGAGCAACTATCTTCTATGGCGCGCCCAACTTGTTCCGTTCCTCAGGAGCGCG
AAACTCATGGGTTATCTTGATGGCACATTGCCGGCGCCTGCCCAACTGATCGCCTCTTCAACGGTGACTG

GTGCTGCGCAAGTGGCAAATCCAGCATATATACGCTGGTATGATCAAGATCAGCAGCTTCTCAGCGGACT
TCTCTCATCGATGACTGAAGATGTTCTTCGGGACGTGGTTGCAGCCACCTCATCCAAGGAGGTGTGGGAC
TCTCTGCAGAAGAAGTTCTCCTCCTCCACTCGAGCTCGCACGGTTCAGATCCGTGTTGAACTCGCTACGGC
AAAGAAGCGTGACTCAACAGCTGCTGACTACTATCGCAAGATTACAGGGCTGGCAAATGAGCTGGCTGC
TGCTGGCGCTCCACTCCAAGATGATGAGGTTCTCTCTTATTTACTTGCTGGCCTTCCTGCTGAATATGATCC
GTTTGTTACTTCCATTACCACCAAGACAGATCCTCTCTCGCTTGACGATGTGTTCGCCCATCTGATGGCGTT
CGAAGCACGACAACTCCAGCATCAGACTGAGATTCAACTAAACACCGGCATGTCCGCGAACTATGTTGG
GCGTGGTGGTCGTGGTCATCGCGGTCGTGGCCGCAACACTCGTGGCCGCGGGCGTTTTGCACCTAGGGG
TGGTGCGCCTCCTCGCGCATCCGGTGGTCATGGCTCATCTTCTCGTCCCACTTGTCAAATTTGTGAAAAGG
AAGGACATACAGCAATCAAGTGCTGGTATCGCATGGATGAATCCTACACTGACACACCTTCATCCTCAGC
TCTGGCGGCCACTACTTCGTACAAGATCGATCCAAACTGGTATCTTGATACGGGCGCTACCGATCACATA
ACCAGCGATCTCGATCGTCTGGCCATACGAGAACGCTACAACGGTGGTGAGCAAGTCCACGTCGGTAAT
GGAGCAGGTTTGCGCATTTTGCATATTGGCCATTCTTCCATTACTACTTCTATGCGTCCACTTGCATTGCGC
AATATTCTTCATGTTCCACAAATTGCTAAGCATCTTCTTTCTGTTCATAAATTCTCTCTTGATAATGATGTAT
ATTTTGAATATCATCCCTGGCATTTTTCCATTAAGGACCGGCACACCAAGAAAAGCCTGCTGGATGGGAA
GTGTGAGGCCGGCCTATATCCCATCAAGCCTTGTGATGTCGTCGCCCTTCGTCAAGCTTTGATTAGTAGG
ACTCCAAGTCGCGCCCAATGGCATGCTCGCCTTGGGCATCCCTCATCTCAAGTGGTCCAGTCAATTCTGCA
TCTTAATAAAATTTTATGTCCTAGTGAGTCTTCCCTTCCAGTTTGTAATGCTTGTCAGTTAGCAAAAAGTCA
TCAATTACCCTATAATAATTCGGTTCATCAGTCGACCTCTCCTTTAGAGATAATTTTTTCCGATGTCTGGGG
TCCTGCGCCCCAATCTGTTAGTGGTTTCAAATATTATATCAGTTTCATCGATGATTTTAGTAAGTTTTGTTG
GATCTATTTGATGCATGATCGTTCTGAAGTCCCTCGTATTTTTCGGCAGTTTCAAACTCATGTCGAACGTCT
TTTAAACACGAAAATTAAATGTGTTCAGTCTGACTGGGGAGGGGAATATCAGAAAATTCATAACACTCTT
TTCCGCCCGCTTGGCATTGCTCACCGTGTTTCTTGTCCACATACACATCAACAAAATGGTTCTGCTGAGCG
TAAACATCGTCACATTTTTGAAACTGGCCTAGCTCTTCTTGCCCATGCACATGTTCCTCTAAAATTTTGGGA
TGAGGCTTTTCTTACTGCTACGTATCTTATCAATCGTCTTCCTACTCGTGTCATAGACAACAAGTGCCCGTT
AGAGCGTCTGCTTAATACTCCTCCAAATTACTCCATGCTCAAAATTTTTGGATGTGCATGCTGGCCGCATC
TTTGAGCTTATAATAATCGCAAACTTGCTTTTCGCTCCAAAGAATGTGTTTTTCTTGGATACAGTTCGCTTC
ACAAAGGGTATAAGTGCCTTGACATAGAATCTGGTCGCATCTATATTTCTCGGGATGTTGTTTTTGATGAA
AATATTTTTCCTTTTGCACGCTCAGCAGCTTCTCATCTTGTCCAATCCCATTCTGATTGTTCCTCCACATTGC
GCAAATTGCAGTCGGACATTAGTAGTACTGATGTGAATTATGATCATGTGCATATGTTGTTGCCTGCTAA
CATTGTTCATGCAGAAAATCCAGCCACCTCATCCCTGCCTGAGTCTCAAGAATGGCTACAGGTTGCAGCG
CCAGCGCCTGCGCTAAATGCGACCACCTCTGCTACAAGAAGGCTCGACGAATCTTCCTCCTTGGCTACATC
TGGTCAGCCTGCTGATACTGGCACATTTACACCAACAGAACATGTGCCTGGGTCATCTGCTTCACCTCAGT
TGCCTTCCGCCGAGGTGCCTGCTGGGGCCACCTTGCCTGCTGATCGTGTACTGCATCAGGATACTGTTGC
TCCTTCGCCTGAATCCCATCCAGATACACATCCATACGGCACCAGGTTGAAACACAACATCAGGATCCCCA
AGAAACGAACAGACGGCACTGTTGCATGGTCAGTTGTTAAGTCCCCTACTGAACCTACAAATTATCGTAT
TGCTATGGAAAATCCTTTTTGGCTCCAGGCTATGAATGATGAATTTCAGGCACTGTTGAAAAATAAGACG
TGGCACTTAATTCCCCCTAATCCAACTCTTAATCTTATTGATTGCAAATGGGTATTTCGGTTAAAGCAGCTT
CCTAATGGATCTATTGATCGCTATAAAGCTCGCCCGGTCGCCAAAGGTTTCACCCAGCAGTATGGTGTGG
ATTATGATGAGACATTTAGTCCGGTTGTTAAGCCCACAACCATTCGTCTATTACTATCTCTTGTTGTTTCTC
GAGGATGGTCTATGCGTCAGATTGACATACAAAATGCTTTTCTTCATGGTATTCTTAATGAAGACGTCTAT
ATGAAGCAACCTCCTGGTTTTCTAGACTCCAAGCATCCGGACTATGTTTGCAAATTGGACAAAGCCTTGTA
TGGCTTGAAGCAATCGCCCCGTGCGTGGTTTTCCCGTCTTAGCTCCAAATTGATACAATTGGGATTTACTG
CTTCCAAGGCCGACGTATCTTTATTCATCTTTAATCAAGATGGTATTCAGATCTACATTCTTATCTACGTTG

ATGATATTATTATTCTCAGTTCATCAACAGTGGCCACTGACAAACTTCTCGTTCAACTCCGGGATGATTTT
GCAGTAAAAGATCTTGGACAGCTTAGCTATTTTCTTGGCATTGAGGTCCATCACACATCTGATGGTATTGT
GTTGCTACAACGCAAATATATTCGTGATCTATTGAGTCGCACCAACATGCTTTCTTCTAACGGTGTGCCCA
CTCCTATGCTTCCCACTGAAAAATTATCTCTTGATGGTGGTGAAAAGCTTTCTTCAGATGATGCGACTAAA
TATCGAAGTGCTGTGGGTGCCTTGCAATATCTTTTATTGACTCGCCCTGACATCTCCTTCAGTGTTAATCG
GGTTTGCCAATTTATGTCTGCTCCAACTGATGTTCATTGGGGAGCTGTAAAACGGATTCTTCGTTATCTTC
ATGCAACTATGGATTTCGGTCTCCGATTTAGTCGCAGGGGATCTTCTCTTCTCAGTGCTTTTTCGGATGCT
GATTGGGCTGGTAATCCGGATGATCGTCGTAGCACCGGTGGCTATGTTATATTTTTTGGTGGGAATCTTG
TTTCTTGGAGTTCACGGAAGCAAGCAACAGTTTCTCGCTCAAGCACTGAATCTGAACATAAAGCTGTTGC
TGATGCTACGGCTGAAGTTATATGGCTCCAAGTTTTATTGCGGGAACTTGGAGTTTCGCAACCACGTTCTC
CTACTCTCTGGTGTGACAATCTTGGTGCTACGTATCTATCGGCCAATCCGATTTTTCATCGTCGCTCCAAGC
ATATTGAAGTTGATTATCACTTTGTCCGTGAACGTGTAGCTACCAAGCAACTTGAAGTACGGCCCATCGCT
ACAAAAGATCAGCTAGCAGATGCATTGACAAAGCCATTGGCGGCACCTGCTTTTATTCGTTTTAGAAACA
ATCTGAACTTAGTATCTCTACGTCCAGATTGAGGGGGGATGTTAAAAGATAGAAGGATATATCCATCCTT
GGGAGCTGCGATCTCCCTGATTTCCTGAACCAATCTATTCTTGGGAAGTTATTATCTCCCTGATTTCTTGG
AGCTACACCTCCGTGATTCTCTAGAGTTGTTACCATGTAAATCCTGTGTACCTCGTTCCTATATATTGAAG
GATGCCGTGACCCCTAGAGGGCAGGCTGTGCATTGCGTCAACACAGAG

>chr9:136263874-136268594

GTTTTTGTTAAAAGATAGAAGGATATATCCATCCTTGGGAGCTGCGATCTCCCTGATTTCCTGAACCAATC
TATTCTTGGGAAGTTATTATCTCCCTGATTTCTTGGAGCTACACCTCCGTGATTCTCTAGAGTTGTTACCAT
GTAAATCCTGTGTACCTCGTTCCTATATATTGAAGGATGCCGTGACCCCTAGAGGGCAGGCTGTGCATTG
CGTCAACATGGTATAGGAGCCAACTATCGATCCCAGATGTCCATAGACTCCTCTCTTTTTTCCGCCTCCTCC
CCTTCCACCATGACTACCAACGCTCCAGCATTCCAATCTGCCTGCGTTCCAATTCCTCTGCACCATGTGGTA
ACCATCCGTCTTACCAAGAGCAACTATCTTCTATGGCGCGCCCAACTTGTTCCGTTCCTCAGGAGCGCGAA
ACTCATGGGTTATCTTGATGGCACATTGCCGGCGCCTGCCCAACTGATCGCCTCTTCAACGGCGACTGGT
GCTGCGCAAGTGGCAAATCCAGCATATATACGCTGGTATGATCAAGATCAGCAGCTTCTCAGCGGACTTC
TCTCATCGATGACTGAAGATGTTCTTCGGGACGTGGTTGCAGCCACCTCATCCAAGGAGGTGTGGGACTC
TCTGCATAAGAAGTTCTCCTCCTCCACTCGAGCTCGCACGGTTCAGATCCGTGTTGAACTCGCTACGGCAA
AGAAGCGTGACTCAACAGCTGCTGACTACTATCGCAAGATTACAGGGCTGGCAAATGAGCTGGCTGTTG
CTGGCGCTCCACTCCAAGATGATGAGGTTCTCTCTTATTTACTTGCTGGCCTTCCTGCTGAATATGATCCGT
TTGTTACTTCCATTACCACCAAGACAGATCCTCTCTCGCTTGACGATGTGTTCGCCCATCTGATGGCGTTC
GAAGCACGACAACTCCAGCATCAGACTGAGATTCAACTAAACACCGGCATGTCCGCGAACTATGTTGGG
CGTGGTGGTCGTGGTCATCGCGGTCGTGGCCGCAACACTCGTGGCCGCGGGCGTTTTGCACCTAGGGGG
GTGCGCCTCCTCGCGCATCCGGTGGTCATGGCTCATCTTCTCGTCCCACTTGTCAAATTTGTGAAAAGGAA
GGACATACAACAATCAAGTGCTGGTATCGCATGGATGAATCCTACACTGAGACACCTTCATCCTCAGCTC
TGGCGGCCACTACTTCGTACAAGATCGATCCAAACTGGTATCTTGATACGGGCGCTACCGATCACATAAC
CAGCGATCTCGATCGTCTGGCCATACGAGAACGCTACAACGGTGGTGAGCAAGTCCACGTTGGTAATGG
AGCAGGTTTGCGCATTTTGCATATTGGCCATTCTTCCATTACTACTTCTATGCGTCCACTTGCATTGCGCAA
TATTCTTCATGTTCCACAAATTGCTAAGCATCTTCTTTCTGTTCATAAATTCTCTCTTGATAATGATGTATAT
TTTGAATATCATCCCTGGCATTTTTCCATTAAGGACCGGCACACCAAGAAAAGCCTGCTGGACGGGAAGT
GTGAGGCCGGCCTATATCCCATCAAGCCTTGTGATGTCGTCGCCCCTTCATCAAGCTTTGATTAGTAGGACT
CCAAGTCGCGCCCAATGGCATGCTCGCCTTGGGCATCCCTCATCTCAAGTGGTCCAGTCAATTCTGCATCT

TAATAAAATTTTATGTCCTAGTGAGTCTTCCCTTCCAGTTTGTAATGCTTGTCAGTTAGCAAAAAGTCATCA
ATTACCCTATAATAATTCGGTTCATCAGTCGACCTCTCCTTTAGAGATGATTTTTTCCGATGTCTGGGGTCC
TGCGCCCCAATCTGTTAGTGGTTTCAAATATTATATCAGTTTCATCGATGATTTTAGTAAGTTTTGTTGGAT
CTATTTGATGCATGATCGTTCTGAAGTCCCTCGTATTTTTCGGCAGTTTCAAACTCATGTCGAACGTCTTTT
AAACACGAAAATTAAATGTGTTCAGTCTGACTGGGGAGGGGAATATCAGAAAATTCATAACACTCTTTTC
CGCCCGCTTGGCATTGCTCACCGTGTTTCTTGTCCACATACACATCAACAAAATGGTTCTGCTGAGCGTAA
ACATCGTCACATTGTTGAAACTGGCCTAGCTCTTCTTGCCCATGCACATGTTCCTCTAAAATTTTGGGATG
AGGCTTTTCTTACTGCTACGTATCTTATCAATCGTCTTCCTACTCGTGTCATAGACAACAAGTGCCCGTTAG
AGCGTCTGCTTAATACTCCTCCAAATTACTCCATGCTCAAAATTTTTGGATGTGCATGCTGGCCGCATCTTC
GAGCTTATAATAATCGCAAACTTGCTTTTCGCTCCAAAGAATGTGTTTTTCTTGGATACAGTTCGCTTCACA
AAGGGTATAAGTGCCTTGACATAGAATCTGGTCGCATCTATATTTCTCGGGATGTTGTTTTTGATGAAAAT
ATTTTTCCTTTTGCACGCTCAGCAGCTTCTCATCTTGTCCAATCCCATTCCGATTGTTCCTCCACATTGCGCA
AATTGCAGTCGGACATTAGTAGTACTGATGTGAATTATGATCATGTGCATATGTCGTTGCCTAATGCCTGC
TAACATTGTTCATGCAGAAAATCCAGCCACCTCATCCCTGCCTGAGTCTCAAGAATGGCTACAGGTTGCA
GCGCCAGCGCCTGCGCTAAATGCGACCACCTCTGCTACAAGAAGGCTCGGCGAATCTTCCTCCTTGGCTA
CATCTGGTCAGCCTGCTGATACTGGCACATTTACACCAACAGAACATGTGCCTGGGTCATCTGCTTCACCT
CAGTTGCCTTCCGCCGAGGTGCCTGCTGGGACCACCTTGCCTGCTGATCGTGTACTGCATCAGGATACTG
TTGCTCCTTCGCCTGAATCCCATCCAGATACACATCCATACGGCACCAGGTTGAAACACAACATCAGGATC
CCCAAGAAACGAACAGACGACACTGTTGCATGGTCAGTTGTTAAGTCCCCTACTGAACCTACAAATTATC
GTATTGCTATGGAAAATCCTTTTTGGCTCCAGGCTATGAATGATGAATTTCAGGCACTGTTGAAAAATAA
GACGTGGCACTTAATTCCCCCTAATCCAACTCTTAATCTTATTGATTGCAAATGGGTATTTCGGTTAAAGC
AGCTTCCTAATGGATCTATTGATCGCTATAAAGCTCGCCTGGTCGCCAAAGGTTTCACCCAGCAGTATGG
TGTGGATTATGATGAGACATTTAGTCCGGTTGTTAAGCCCACAACCATTCGTCTATTACTATCTCTTGCTAT
TTCTCGAGGATGGTCTATGCGTCAGATTGACATACAAAATGCTTTTCTTCATGGTATTCTTAATGAAGACG
TCTATATGAAGCAACCTCCTGGTTTTCGACTCCAAGCATCCGGACTATGTTTGCAAATTGGACAAAGCCTT
GTATGGCTTGAAGCAATCGCCCCGTGCGTGGTTTTCCCGTCTTAGCTCCAAATTGATACAATTGGGATTTA
CTGCTTCCAAGGCCGACGTATCTTTATTCATCTTTAATCAAGATGGTATTCAGATCTACATTCTTATCTACG
TTGATGATATTATTATTCTCAGTTCATCAACAGTGGCCACTGACAAACTTCTCGTTCAACTCCGGGATGAT
TTTGCAGTAAAAGATCTTGGACAGCTTAGCTATTTTCTTGGCATTGAGGTCCATCACACATCTGATGGTAT
TGTGTTGCTACAGCGCAAATATATTCGTGATCTATTGAGTCGCACCAACATGCTGTCTTCTAACGGTGTGC
CCACTCCTATGCTTCCCACTGAAAAATTATCTCTTGATGGTGGTGAAAAGCTTTCTTCAGATGATGCGACT
AAATATCGAAGTGCTGTGGGTGCCTTGCAATATCTTTTATTGACTCGCCCTGACATCTCCTTCAGTGTTAA
TCGGGTTTGCCAATTTATGTCTGCTCCAACTGATGTTCATTGGGGAGCTGTAAAACGGATTCTTCGTTATC
TTCATGCAACTATGGATTTCGGTCTCCGATTTAGTCGCAGGGGATCTTCTCTTCTCAGTGCTTTTCGGATG
CTGATTGGGCTGGTAATCCGGATGATCGTCGTAGCACCGGTGGCTATGTTATATTTTTTGGTGGGAATCT
TGTTTCTTGGAGTTCACGGAAGCAAGCAACAGTTTCTCGCTCAAGCACTGAATCTGAATATAAAGCTGTT
GCTGATGCTACGGCTGAAGTTATATGGCTCCAAGTTTTATTGCGGGAACTTGGAGTTTCGCAACCACGTT
CTCCTACTCTCTGGTGTGACAATCTTGGTGCTACGTATCTATCGGCCAATCCGATTTTTCATCGTCGCTCCA
AGCATATTGAAGTTGATTATCACTGCATTGACAAAGCCATTGGCGGCACCTGCTTTTATTCGTTTTAGAAA
CAATCTGAACTTAGTATCTCTACGTCCAGATTGAGGGGGGATGTTAAAAGATAGAAGGATATATCCATCC
TTGGGAGCTGCGATCTCCCTGATTTCCTGAACCAATCTATTCTTGGGAAGTTATTATCTCCCTGATTTCTTG
GAGCTACACCTCCGTGATTCTCTAGAGTTGTTACCATGTAAATCCTGTGTACCTCGTTCCTATATATTGAA
GGATGCCGTGACCCCTAGAGGGCAGGCTGTGCATTGCGTCAACAGTTCT

NC358

>chr5:121493746-121497662

TGGTTCGATTGCTCCATGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTTGCGTGGAGT
TGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATAGTGGGCCGG
CCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAGAAACAGAAT
CCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGGCGCGGCTCA
CGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCATCTGGTATCGGATTCGGGGTCGGCTGCGT
CTTCATCCTCTTCCACGTGCTGTTCCTCGTCTTTGGCGGTTCGTGGGCGACTTCTTCTTCTGCTGGGTCCAT
CCAGTTCATCACTACGGCGCCCGCAACCAAACCATCGGTACAGACTCGTTTGGTGTTACAGGCTATGGCG
GATGGGGAGCGAGTTAGTCAGGAGAATTGGAGTCGTGTTATTGAGACGTTGGAGGGAGTGAGTAGCAG
GTTGAAGGATGTGGAGAGGGTGCAGCAACAAATCATGATGCGGACTGAATTAGCGGCATCAGTTGCTG
AACAGGCTGCGAAGGAGAGAAATGATTTGGGTCGTAAATTGGAGGAGACTGGAAGGGGAATGATATTT
GGGCTTTAGTGGTGTTGGAAGAGGTGTCTGCAGATCAGGGAGTACAGCAGCAGGGGGATCTTCAGCAG
TTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTACCTCCTGCTAGGGTGTATGATCATCA
TATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTATAGGTACTCTCCTCTTCACAAGGATG
AAATTGAGAGACAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATTGTTCCTAGTGTAAGCCCGTTTGCTTCT
CCGGTATTGTTGGTTAAGAAGAAAGATAACACTTGGAGGTTTTGTGTGGACTACAGGAAATTAAATTCTT
TGACCATTAAGAACAGATTTCCTATGCCGCTTGTAGAAGAGATTCTGGATGAACTGGCTGGTACTAAATA
TTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAATGGGGGAAGGTGATGAGCATAAAACA
GCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCCTTTTGGGTTAACCAATGCTCCTGCAAC
CTTTCAATGTGCTATGAATTCTGTTTTGGCTCCTTTTCTTCGGAAATTTGTCATGGTTTTCATTGATGACATT
TTGGTCTATAGTGCTACATGGGAGGAACACTTGGCTCATATTAAGATGGTGCTGCAGACTTTGAGAGAG
CACCAGTTCTATCTGAAGAGGAGTAAATGTGTGTTTGGGAGGAATGAGCTCATTTATTTGGGACATATTA
TCTCTCACACAGGAGTGGCCACTGATCCTGCTAAGACAACAGCCATGGAGAAATGGCCTGTTCCTCAGTC
AGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTATAGGAGATTTGTGAGACATTATGGGAT
TATTGCAAAACCTCTTACTAATCTGTTGAAGCATAAGGCTTTTCAGTGGAGTCCTGAGGCTGATCAGGCTT
TTCAAGCGCTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCCCTACCCAGGTTTGATGAGTTATTTGTT
GTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGATGCAATGTGATAGACCAGTGGCATAT
CTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTATCCATTTATGAAAAAGAGTTTTTGGCTCTAATCA
TTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATTTGTGATTAAAACTGATCATCAGGCTT
TGAGTTTTCTGGATGATCAGCTTTTGCAATCTGATATGCAGAGGAAGGCTATGGCTAAGTTGATGGGGCT
TCATTTCAAAATTGTTTACAAAAAAGGCAAGGACAATGTGGCTGCCGATGCTCTATCCAGGATTGGTCAT
GTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGGGTGCAAGAACTCCTGAATGCTTATGTG
ACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGGCCAGTCCTAATGAGCAAGGTTTTTCTT
TACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGCTAACAATTCTGCCCTTCAAACTAAGTT
GATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGTGGTGAGGCCACTTACCAAAGGATCAA
GAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAGTTTGTGCAGCAATGTGAGGTGTGTCA
GCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAACCTTTACCAGTGCCTCAAGGGGCTTGG
CAGGATCTTACATTGGATTTCATTGAAGGTCTACCTAGATCTGAGGGATATGATACTATTTTGGTCGTGGT
GGATAGATATTCCAAATATGCTCATTTTTTTCCCCTCAAGCATCCTTTCACTGCTGTTGGAGTGGCACAGG
TGTTTCTTGACAATATTGTCAAGTTGCATGGGACACCAAAGTCTTTGGTTTCCGATAGAGACGAAGTGTTC
ACTAGTTTGTTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCAAATTGGCTCTCACTACTGCATATCACCCT
CAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGAAATGTATCTTCGCTGTTCTATTCATGAGG
CACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAATTTTGGTACAATACTTCATATCACACAGCTTTG
GGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTGTTTGCTGCTGCTCCTATGCTTTCTGGGGT

TGCTCCTAGTTCTGTGACTGAGGTGTTGGCTGATAGACAGTTACATGCTAGTATGCTGAAGGAGAAGTTG
GCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCATAGATCTGACAGAGAGTTTCAGATTGGA
GATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCGGTGGTTAGTCGACCCTGTCCTAAATTGG
CCTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGGGATGGTTGCTTATAAGTTGGATTTACCA
GTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGTTGAAGGCTTTTACTCCCAATTATTCTCCGGTG
TTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGATTATGCCATTGGCTGTGGTGGAAAGGAG
GATGGTGAAGAAGGGTAATCATCCTGTGGTGCAAGTAAGGGTGACTTGGTCCAATATGCCAGCTGAGGT
TACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTTCCTACAGCTCTTGATTGGGGACAATCAAGC
TCTTCCGGGGGGGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGT
TTTGCGTGGAGTTGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAA
TAGTGGGCCGGCCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAAT
AGAAACAGAATCCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTT
GGCGCGGCTCACGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCACTCCAGTGATGATGCATG
GT

>chr1:281711216-281715346

GACCATGTTAGTACTGCTACTGGTATGGCAATTTAATCTTTGTGCTGATTTGCGTGGAGTTGTAAATGGG
CTGTTAAGCCAGGATTTGGGGTGTCAATGTGGCTAATTAGCCTTGTAATAGTGGGCCGGCCTTATTAGCT
TGACTATAAAAGCTGTAATCTGGTGTGGATCTGGGGCCTTAAGTATGAATTGCAGGATGTGGTGCAGGT
TCAAACACCTTCTACGGTGGACCGTGCTATTCATTTGGCTCAGCTTCAACAGGAGGTGTTGGAGCGCAAT
AAATTTCGGGGACAACGGTCAGGGTTGTCTATGCATGTGAAACAACCTGTGGCTGGCCAGAAAACAGAT
GACCAAGGACTCGGGGCAGATCTCACAAGAGCACGACAACTTAGGGATTTCAGGAGACTTAATGGGTTA
TGTTATGCTTGTGGGGAAAAGTTTGAGCCTGGTCATATTGCCAAATGTGCAAAGCGTGGAGCACTGCAG
TTAAATGTAGTGGCTACTGAGGATATGTCTGAGATACTGACTGATGATACATTACAGCAGTTGGCACAAG
AAGATGAAGTTAAGGAGTTGTGTTGTCGTTTATCTGTGCAGGCTCTAGCTGGTACAGGCAGTATTAACAG
TATGAGGATTCGAGCGGTAGTGCAGAAACAGTTGATGTTGATGTTGGTGGACTCAGGCAGTTCGACCAG
TTTTGTGAGTAGTCATGTGGTGCATCAGTTGGGATTGACTGTGCAAGTTTGTCCAGCTGTGACAGTCAAA
GTGGCAAATGGAGAGACTATGGTCAGTAATACTATGGTGAAAGGATTGGAATGGTGGGCTGGTGCACA
ATTGTTTCATACGGATATGAGGGTTTTAACTTTGAGTGCTTTTGATGCTATTTTGGGGTATGATTGGTTGA
GATACCATAGCCCTATGGAATGTGACTGGGAACAAAAGTTTTTGAAGTTTACAATAGGGGGCAAAGAGG
TAGTGTTGCAGGGTGATGAGGAGGCTGGGGATGCGAAGGTGAAGCCTGTGACAGGTTCCCAGTTGCAC
AAATGGGTGAAGGGGCATGATATTTGGGCTTTAGTGGTGTTGGAAGAAGTGTCTGAATATCAGGGAGT
ACAGCAGCAGGGGGATCTTCAGCAGTTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTA
CCTCCTGCTAGGGTGTATGATCATCATATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTA
TAGGTATTCTCCTCTTCACAAGGATGAAATTGAGAAGCAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATT
GTTCCTAGTGTAAGCCCGTTTGCTTCTCCGGTATTGTTGGTTAAGAAGAAAGATAGCACTTGGAGGTTTT
GTGTGGACTACAGGAAATTAAATTCTTTGACCATTAAGAACAGATTTCCTATGCCTCTTGTTGAAGAGATT
CTGGATGAACTGGCTGGTACTAAATATTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAAT
GGGGGAAGGTGATGAGCATAAAACAGCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCC
TTTTGGGTTAACCAATGCTCCTGCAACCTTTCAATGTGCGATGAATTCTGTTTTGGCTCCTTTTCTTCGGAA
ATTTGTCATGGTTTTCATTGATGACATTTTGGTTTATAGTGATACATGGGAGGCACACTTGGCTCATATTA
AGATGGTGCTGCAGACTTTGAGAGAGCACCAGTTTTATCTGAAGAGGAGCAAATGTGTGTTTGGGAGGA
ATGAGCTCATTTATTTGGGACATATTATTTCTCACACAGGGGTGGCCACTGATCCTGCTAAGACAGCAGC

CATGGAGAAATGGCCTGTTCCTCAGTCAGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTAT
AGGAGATTTGTGAGACATTATGGGATTATTGCAAAACCCCTTACTAATCTGTTGAAGCATAAGGCTTTTCA
GTGGAGTCCTGAGGCTGATCAGGCTTTTCAAGCGTTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCC
CTACCCAGGTTTGATGAGTTATTTGTTGTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGA
TGCAATGTGATAGACCAGTGGCATATCTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTCTCCATTTA
TGAAAAAGAGTTTTTGGCTCTAATCATTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATT
TGTGATTAAGACTGATCATCAGGCTTTGAGTTTTCTGGATGATCAGATTTTGCAATCTGATATGCAGAGG
AAGGCTATGGCTAAGTTGATGGGGCTTCATTTCAGAATTGTTTACAAAAAAGGCAAGGACAATGTGCCT
GCCGATGCTCTATCCAGGATTGGTCATGTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGG
GTGCAAGAACTCCTGAATGCTTATGTGACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGG
CCAGTCCTAATGAGCAAGGTTTTTCTTTACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGC
TAACAATTCTGCCCTTCAAACTAAGTTGATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGT
GGTGAGGCCACTTACCAAAGGATCAAGAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAG
TTTGTGCAGCAATGTGAGGTGTGTCAGCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAAC
CTTTACCAGTGCCTCAAGGGGCTTGGCAGGATCTTACATTGGATTTCATTGAAGGTCTCCCTAGATCTGA
GGGATATGATACAATTTTGGTCGTGGTGGATAGATATTCCAAATATGCTCATTTTTTCCCCCTCAAGCATC
CTTTCACTGCTGTTGGAGTGGCACAGGTGTTTCTTGACAATGTTATCAAGTTGCATGGGACACCAAAGTC
TTTGGTTTCCGATAGAGACAAAGTGTTCACTAGTTTATTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCA
AATTGGCTCTTACTACTGCATATCACCCTCAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGA
AATGTATCTTCGCTGTTCTATTCATGAGGCACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAGTTTT
GGTACAATACTTCATATCACACAGCTTTGGGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTG
TTTGCTGCTGCTCCTATGCTTTCTGGGGTTGCTCCTAGTTCTGTGACTGAGGTGTTGGCGGACAGACAGTT
ACATGCTACCATGCTGAAGGAGAAGTTGGCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCA
TAGATCTGACAGAGAGTTTCAGATTGGAGATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCG
GTGGTTAGTCGACCCTGTCCTAAATTGGCTTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGG
GGTGGTTGCTTATAAGCTGGATTTACCAGTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGCTGA
AGGCTTTTACTCCCAATTATTCTCCGGTGTTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGG
TTAAGCCATTGGCTGTGGTGGAAAGGAGGATGGTGAACAAGGGTAATCATCCTGTGGTGCAAGTAAGG
GTGACTTGGTCCAATATGCCAGCTGAGGTTACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTT
CCTACAGCTCTTGATTGGGGACAATCAAGCTCTTCTGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTAT
GGCAATTTAATCTTTGTGCTGATTTGCGTGGAGTTGTAAATGGGCTGTTAAGCCAGGATTTGGGGTGTCA
ATGTGGCTAATTAGCCTTGTAATAGTGGGCCGGCCTTATTAGCTTGACTATAAAAGCTGTAATCTGGTGT
GGATCTGGGGATAAG

Oh43

>chr5:122895255-122899177

TGGTTCGATTGCTCCATGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTTGCGTGGAGT
TGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATAGTGGGCCGG
CCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAGAAACAGAAT
CCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGGCGCGGCTCA
CGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCATCTGGTATCGGATTCGGGGTCGGCTGCGT
CTTCATCCTCTTCCACGTGCTGTTCCTCGTCTTTGGCGGTTCGTGGGCGACTTCTTCTTCTGCTGGGTCCAT
CCAGTTCATCACTACGGCGCCCGCAACCAAACCATCGGTACAGACTCGTTTGGTGTTACAGGCTATGGCG
GATGGGGAGCGAGTTAGTCAGGAGAATTGGAGTCGTGTTATTGAGACGTTGGAGGGAGTGAGTAGCAG

GTTGAAGGATGTGGAGAGGGTGCAGCAACAAATCATGATGCGGACTGAATTAGCGGCATCAGTTGCTG
AACAGGCTGCGAAGGAGAGAAATGATTTGGGTCGTAAATTGGAGGAGACTGGAAGGGGAATGATATTT
GGGCTTTAGTGGTGTTGGAAGAGGTGTCTGCAGATCAGGGAGTACAGCAGCAGGGGGATCTTCAGCAG
TTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTACCTCCTGCTAGGGTGTATGATCATCA
TATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTATAGGTACTCTCCTCTTCACAAGGATG
AAATTGAGAGACAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATTGTTCCTAGTGTAAGCCCGTTTGCTTCT
CCGGTATTGTTGGTTAAGAAGAAAGATAGCACTTGGAGGTTTTGTGTGGACTACAGGAAATTAAATTCTT
TGACCATTAAGAACAGATTTCCTATGCCGCTTGTAGAAGAGATTCTGGATGAACTGGCTGGTACTAAATA
TTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAATGGGGGAAGGTGATGAGCATAAAACA
GCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCCTTTTGGGTTAACCAATGCTCCTGCAAC
CTTTCAATGTGCTATGAATTCTGTTTTGGCTCCTTTTCTTCGGAAATTTGTCATGGTTTTCATTGATGACATT
TTGGTCTATAGTGCTACATGGGAGGAACACTTGGCTCATATTAAGATGGTGCTGCAGACTTTGAGAGAG
CACCAGTTCTATCTGAAGAGGAGTAAATGTGTGTTTGGGAGGAATGAGCTCATTTATTTGGGACATATTA
TCTCTCACACAGGAGTGGCCACTGATCCTGCTAAGACAACAGCCATGGAGAAATGGCCTGTTCCTCAGTC
AGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTATAGGAGATTTGTGAGACATTATGGGAT
TATTGCAAAACCTCTTACTAATCTGTTGAAGCATAAGGCTTTTCAGTGGAGTCCTGAGGCTGATCAGGCTT
TTCAAGCGCTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCCCTACCCAGGTTTGATGAGTTATTTGTT
GTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGATGCAATGTGATAGACCAGTGGCATAT
CTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTATCCATTTATGAAAAAGAGTTTTTGGCTCTAATCA
TTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATTTGTGATTAAAACTGATCATCAGGCTT
TGAGTTTTCTGGATGATCAGCTTTTGCAATCTGATATGCAGAGGAAGGCTATGGCTAAGTTGATGGGGCT
TCATTTCAAAATTGTTTACAAAAAAGGCAAGGACAATGTGGCTGCCGATGCTCTATCCAGGATTGGTCAT
GTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGGGTGCAAGAACTCCTGAATGCTTATGTG
ACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGGCCAGTCCTAATGAGCAAGGTTTTTCTT
TACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGCTAACAATTCTGCCCTTCAAACTAAGTT
GATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGTGGTGAGGCCACTTACCAAAGGATCAA
GAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAGTTTGTGCAGCAATGTGAGGTGTGTCA
GCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAACCTTTACCAGTGCCTCAAGGGGCTTGG
CAGGATCTTACATTGGATTTCATTGAAGGTCTACCTAGATCTGAGGGATATGATACTATTTTGGTCGTGGT
GGATAGATATTCCAAATATGCTCATTTTTTTCCCCTCAAGCATCCTTTCACTGCTGTTGGAGTGGCACAGG
TGTTTCTTGACAATATTGTCAAGTTGCATGGGACACCAAAGTCTTTGGTTTCCGATAGAGACGAAGTGTTC
ACTAGTTTGTTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCAAATTGGCTCTCACTACTGCATATCACCCT
CAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGAAATGTATCTTCGCTGTTCTATTCATGAGG
CACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAATTTTGGTACAATACTTCATATCACACAGCTTTG
GGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTGTTTGCTGCTGCTCCTATGCTTTCTGGGGT
TGCTCCTAGTTCTGTGACTGAGGTGTTGGCTGATAGACAGTTACATGCTAGTATGCTGAAGGAGAAGTTG
GCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCATAGATCTGACAGAGAGTTTCAGATTGGA
GATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCGGTGGTTAGTCGACCCTGTCCTAAATTGG
CCTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGGGATGGTTGCTTATAAGTTGGATTTACCA
GTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGTTGAAGGCTTTTACTCCCAATTATTCTCCGGTG
TTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGATTATGCCATTGGCTGTGGTGGAAAGGAG
GATGGTGAAGAAGGGTAATCATCCTGTGGTGCAAGTAAGGGTGACTTGGTCCAATATGCCAGCTGAGGT
TACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTTCCTACAGCTCTTGATTGGGGACAATCAAGC
TCTTCCGGGGGGGGGGGGGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTATGGCGATTTAATCTTT

GTGCTGTTTTGCGTGGAGTTGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGC
TTTGTAATAGTGGGCCGGCCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAA
AGAAATAGAAACAGAATCCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCT
GGATTTGGCGCGGCTCACGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCACTCCAGTGATGA
TGCATGGT

PE0075

>chr4 Zea mays cultivar PE0075 chromosome 4 whole genome shotgun sequence:165404143-165408930

CAGAGTGTTAAAAGATAGAAGGATATATCCATCCTTGGGAGCTGCGATCTCCCTGATTTCCTGAACCAAT
CTATTCTTGGGAAGTTATTATCTCCCTGATTTCTTGGAGCTACACCTCCGTGATTCTCTAGAGTTGTTACCA
TGTAAATCCTGTGTACCTCGTTCCTATATATTGAAGGATGCCGTGACCCCTAGAGGGCAGGCTGTGCATT
GCGTCAACATGGTATAGGAGCCAACTATCGATCACAGATGTCCACAGACTCCTCTCTTTTTTCCGCCTCCT
CCCCTTCCACCATGACTACCAACGCTCCAGCATTCCAATCTGCCTGCGTTCCAATTCCTCTGCACCATGTGG
TAACCATCCGTCTTACCAAGAGCAACTATCTTCTATGGCGCGCCCAACTTGTTCCGTTCCTCAGGAGCGCG
AAACTCATGGGTTATCTTGATGGCACATTGCCGGCGCCTGCCCAACTGATCGCCTCTTCAACGGTGACTG
GTGCTGCGCAAGTGGCAAATCCAGCATATATACGCTGGTATGATCAAGATCAGCAGCTTCTCAGCGGACT
TCTCTCATCGATGACTGAAGATGTTCTTCGGGACGTGGTTGCAGCCACCTCATCCAAGGAGGTGTGGGAC
TCTCTGCAGAAGAAGTTCTCCTCCTCCACTCGAGCTCGCACGGTTCAGATCCGTGTTGAACTCGCTACGGC
AAAGAAGCGTGACTCAACAGCTGCTGACTACTATCGCAAGATTACAGGGCTGGCAAATGAGCTGGCTGC
TGCTGGCGCTCCACTCCAAGATGATGAGGTTCTCTCTTATTTACTTGCTGGCCTTCCTGCTGAATATGATCC
GTTTGTTACTTCCATTACCACCAAGACAGATCCTCTCTCGCTTGACGATGTGTTCGCCCATCTGATGGCGTT
CGAAGCACGACAACTCCAGCATCAGACTGAGATTCAACTAAACACCGGCATGTCCGCGAACTATGTTGG
GCGTGGTGGTCGTGGTCATCGCGGTCGTGGCCGCAACACTCGTGGCCGCGGGCGTTTTGCACCTAGGGG
TGGTGCGCCTCCTCGCGCATCCGGTGGTCATGGCTCATCTTCTCGTCCCACTTGTCAAATTTGTGAAAAGG
AAGGACATACAGCAATCAAGTGCTGGTATCGCATGGATGAATCCTACACTGACACACCTTCATCCTCAGC
TCTGGCGGCCACTACTTCGTACAAGATCGATCCAAACTGGTATCTTGATACGGGCGCTACCGATCACATA
ACCAGCGATCTCGATCGTCTGGCCATACGAGAACGCTACAACGGTGGTGAGCAAGTCCACGTCGGTAAT
GGAGCAGGTTTGCGCATTTTGCATATTGGCCATTCTTCCATTACTACTTCTATGCGTCCACTTGCATTGCGC
AATATTCTTCATGTTCCACAAATTGCTAAGCATCTTCTTTCTGTTCATAAATTCTCTCTTGATAATGATGTAT
ATTTTGAATATCATCCCTGGCATTTTTCCATTAAGGACCGGCACACCAAGAAAAGCCTGCTGGATGGGAA
GTGTGAGGCCGGCCTATATCCCATCAAGCCTTGTGATGTCGTCGCCCTTCGTCAAGCTTTGATTAGTAGG
ACTCCAAGTCGCGCCCAATGGCATGCTCGCCTTGGGCATCCCTCATCTCAAGTGGTCCAGTCAATTCTGCA
TCTTAATAAAATTTTATGTCCTAGTGAGTCTTCCCTTCCAGTTTGTAATGCTTGTCAGTTAGCAAAAAGTCA
TCAATTACCCTATAATAATTCGGTTCATCAGTCGACCTCTCCTTTAGAGATAATTTTTTCCGATGTCTGGGG
TCCTGCGCCCCAATCTGTTAGTGGTTTCAAATATTATATCAGTTTCATCGATGATTTTAGTAAGTTTTGTTG
GATCTATTTGATGCATGATCGTTCTGAAGTCCCTCGTATTTTTCGGCAGTTTCAAACTCATGTCGAACGTCT
TTTAAACACGAAAATTAAATGTGTTCAGTCTGACTGGGGAGGGGAATATCAGAAAATTCATAACACTCTT
TTCCGCCCGCTTGGCATTGCTCACCGTGTTTCTTGTCCACATACACATCAACAAAATGGTTCTGCTGAGCG
TAAACATCGTCACATTTTTGAAACTGGCCTAGCTCTTCTTGCCCATGCACATGTTCCTCTAAAATTTTGGGA
TGAGGCTTTTCTTACTGCTACGTATCTTATCAATCGTCTTCCTACTCGTGTCATAGACAACAAGTGCCCGTT
AGAGCGTCTGCTTAATACTCCTCCAAATTACTCCATGCTCAAAATTTTTGGATGTGCATGCTGGCCGCATC
TTTGAGCTTATAATAATCGCAAACTTGCTTTTCGCTCCAAAGAATGTGTTTTTCTTGGATACAGTTCGCTTC
ACAAAGGGTATAAGTGCCTTGACATAGAATCTGGTCGCATCTATATTTCTCGGGATGTTGTTTTTGATGAA

AATATTTTTCCTTTTGCACGCTCAGCAGCTTCTCATCTTGTCCAATCCCATTCTGATTGTTCCTCCACATTGC
GCAAATTGCAGTCGGACATTAGTAGTACTGATGTGAATTATGATCATGTGCATATGTTGTTGCCTGCTAA
CATTGTTCATGCAGAAAATCCAGCCACCTCATCCCTGCCTGAGTCTCAAGAATGGCTACAGGTTGCAGCG
CCAGCGCCTGCGCTAAATGCGACCACCTCTGCTACAAGAAGgctcgacgaatcttcctccttggCTACATCTGGTC
AGCCTGCTGATACTGGCACATTTACACCAACAGAACATGTGCCTGGGTCATCTGCTTCACCTCAGTTGCCT
TCCGCCGAGGTGCCTGCTGGGGCCACCTTGCCTGCTGATCGTGTACTGCATCAGGATACTGTTGCTCCTTC
GCCTGAATCCCATCCAGATACACATCCATACGGCACCAGGTTGAAACACAACATCAGGATCCCCAAGAAA
CGAACAGACGGCACTGTTGCATGGTCAGTTGTTAAGTCCCCTACTGAACCTACAAATTATCGTATTGCTAT
GGAAAATCCTTTTTGGCTCCAGGCTATGAATGATGAATTTCAGGCACTGTTGAAAAATAAGACGTGGCAC
TTAATTCCCCCTAATCCAACTCTTAATCTTATTGATTGCAAATGGGTATTTCGGTTAAAGCAGCTTCCTAAT
GGATCTATTGATCGCTATAAAGCTCGCCCGGTCGCCAAAGGTTTCACCCAGCAGTATGGTGTGGATTATG
ATGAGACATTTAGTCCGGTTGTTAAGCCCACAACCATTCGTCTATTACTATCTCTTGTTGTTTCTCGAGGAT
GGTCTATGCGTCAGATTGACATACAAAATGCTTTTCTTCATGGTATTCTTAATGAAGACGTCTATATGAAG
CAACCTCCTGGTTTTCTAGACTCCAAGCATCCGGACTATGTTTGCAAATTGGACAAAGCCTTGTATGGCTT
GAAGCAATCGCCCCGTGCGTGGTTTTCCCGTCTTAGCTCCAAATTGATACAATTGGGATTTACTGCTTCCA
AGGCCGACGTATCTTTATTCATCTTTAATCAAGATGGTATTCAGATCTACATTCttatctacgttgatgatattatta
ttCTCAGTTCATCAACAGTGGCCACTGACAAACTTCTCGTTCAACTCCGGGATGATTTTGCAGTAAAAGAT
CTTGGACAGCTTAGCTATTTTCTTGGCATTGAGGTCCATCACACATCTGATGGTATTGTGTTGCTACAACG
CAAATATATTCGTGATCTATTGAGTCGCACCAACATGCTTTCTTCTAACGGTGTGCCCACTCCTATGCTTCC
CACTGAAAAATTATCTCTTGATGGTGGTGAAAAGCTTTCTTCAGATgatgcgactaaatatcgaagtgcTGTGGG
TGCCTTGCAATATCTTTTATTGACTCGCCCTGACATCTCCTTCAGTGTTAATCGGGTTTGCCAATTTATGTC
TGCTCCAACTGATGTTCATTGGGGAGCTGTAAAACGGATTCTTCGTTATCTTCATGCAACTATGGATTTCG
GTCTCCGATTTAGTCGCAGGGGATCTTCTCTTCTCAGTGCTttttcggatgctgattgggctggTAATCCGGATGAT
CGTCGTAGCACCGGTGGCTATGTTATATTTTTTGGTGGGAATCTTGTTTCTTGGAGTTCACGGAAGCAAG
CAACAGTTTCTCGCTCAAGCACTGAATCTGAACATAAAGCTGTTGCTGATGCTACGGCTGAAGTTATATG
GCTCCAAGTTTTATTGCGGGAACTTGGAGTTTCGCAACCACGTTCTCCTACTCTCTGGTGTGACAATCTTG
GTGCTACGTATCTATCGGCCAATCCGATTTTTCATCGTCGCTCCAAGCATATTGAAGTTGATTATCACTTTG
TCCGTGAACGTGTAGCTACCAAGCAACTTGAAGTACGGCCCATCGCTACAAAAGATCAGCTAGCAGATG
CATTGACAAAGCCATTGGCGGCACCTGCTTTTATTCGTTTTAGAAACAATCTGAACTTAGTATCTCTACGT
CCAGATTGAGGGGGGATGTTAAAAGATAGAAGGATATATCCATCCTTGGGAGCTGCGATCTCCCTGATTT
CCTGAACCAATCTATTCTTGGGAAGTTATTATCTCCCTGATTTCTTGGAGCTACACCTCCGTGATTCTCTAG
AGTTGTTACCATGTAAATCCTGTGTACCTCGTTCCTATATATTGAAGGATGCCGTGACCCCTAGAGGGCA
GGCTGTGCATTGCGTCAACACAGAG

Tx303

>chr9:21661623-21666549

CCTCCTGTTACCAGCCAGGCGTCAGTATCTGATCACCCACAGAAGCAGCCCCAGCGTCGCAAGCCTGGAA
CGCGTGAGCGTCGTGCCAATGTCCGCGTCACAGGACCGGAGTGGGTGTAGTTGTGTGCGTGGTGTTCCT
GTCCCTGTTTGATGGACCAGCTGCGTCTGGTGGAGGGTTTCTTTATGTATCGCTAACTCTGAACTGGGAA
GGCAGGAAACAAACAGAGAGAAAATATCTATTCCGGCCAGTTCCTATCTCTCTCTCTGTCGTCCTCCTCCT
CCAACCTTTTCCCCATTTCCTTCTGTTAGTATAGATCAGAGCTGCTAACAAATTGGTATCAGAATCCTCCGT
TCCTCGGCCGTCGCGTCCATGGCGTCTCCTGTCGACGCGCAGCTCACGCAGGCGCTGGAACAACAGACCC
GCGCCATTCTCGCTGAGCTCGACAAACGGTTCACAGCGATCGACTCCAAGTGGGAACTCCGCGTCGGCG
CGCTGGAGTCCCAGGCGGCGGACTTCGCTCGCCACTCCGACGAGTTCGCCTCTGCCATCCGCGCGGACG

TCGAAGCTCACCTTTCGGCGGCCGATGCGGATCTTACCGATCGCGTCCGCTACTGGGAGGCCACTACCGG
AGCGCGCGTCACCTCCCTCGAGTCCGCCCAGCACACGTTCGACGCTTGGCGCCCCAGGATGGAATCCTCC
GTCGACGCCCTACAATCTTCCATGGATTCCGTTCGCGCCGAGCTATCCAAGGTGGAGATCCGGTGGTCCC
AAGGGGCGCGCGTCGAGGGCTTCTCCAAGCCCGGCCTCCTCGGCTCGCCCAGATCGGCGCCGGTGTGTT
CTTCTGCCACCGGCGGTCAAGCTGACGGCTCACGTTTTGGGCCCCGCTTCGCTTCATCTCACCGGGAATG
TGGATTTGGGCCCCCTGTGGCCGATCTCCACTACCCGGTCAAGGGTACGTGGTCACCTCACGCCCCTGCT
TCGTTTGAGTTTCCAGTTCATCATAATCTTAATTCTGGGGAGCTAGATCGTCGGGGGGAACTGAGTAGAG
GATTGGGTGGTCTTCCTAAGGTCAATTTCCCTGTGTTCGATGGCAGTAATCCCAAACTTTGGCAAAAACA
GTGCGAAGATTATTTTGACATGTACGCTACCGAGGAGTTCGTTTGGATTAAAGCCGCAACTATGCATTTC
CATGGCGTAGCTGCGCGCTGGCTCCAATCCATCGAACCTCGCCTGCCCTCGCTCTCCTGGCGACAGTTCT
GTCAGTTCGTCCATGACAGATTTGGTCGTCAACAACATGAGCTAGTTATCCAAAAATTATTCCACATCAAA
CAGACCTCCACTGTCCAGGATTACATTGATCGCTTTTGCGAGTTAGTGGATTTGCTTGTCACCTATGAACA
CACAACAGACCCATTGTATTACACTATGAGATTTATTGATGGCTTGCGTGCTGAAATCAAGTCTGTCATTC
TTGTACAACGTCCTGGCGACCTCGATACTGCCTGCTCTTTGGCCTTGTTGCAGGAAGCTGCAGAGTCTTCT
CGTCGGCCGGAATTGGTTCGGGGTGACCATTCTTCGTTTTCTCGGCCGCTGCTAAGAGCCGCTCCCACGT
CGTCCTACCGTGGGGAGCAGCCTCGTGGTGGCGGGGGCGATTATACCAAGACCAACACGACGGGGCCT
ACAACTGTTGATTCTAAGGTGGCCTCTCTTCGTGCGTACCGGCGTGCTAAGGGTCTCTGTCAATACTGCG
CAGAGAAATGGAACAAGGGTCATAAATGTGCAGATACAGTGCAGTTGCATGCTGTCCAAGAATTGTGGG
AGATGCTGTCAACTGATCAGCCTGAATCTGAAGGAGAGTTTGAGGATTCTGCTGAACAATTTATGATATT
ATTGTCTACAGAAGCTGCCTCGGTTCACCGCCCATCAAAATCATTCAGAATTCGAGGTGTCCTGCAAGGG
GTTGACCTGCTCATGTTGATCGACTCAGGTAGTTCTCATAGCTTCCTGAATTCTGCCCATACTTCTAAGTTG
TCTGGCATCACTTCTATGGATACACCATTGGCTGTACGAGTAGCAAATGGTACTGTGTTGCAGTGTTCTGT
GGAACTCCCTCAGGCTTCTTGGTCCGTGCAGGATCTTAGCTTCTGTACTACATTCAAGTTGATCCATCTAC
CCTTCTATGATGCAATATTGGGTATGGATTGGCTGGAACAGCTGAGCCCAATGACAGTGGACTGGAAAC
ACAAATGGTTGTCCATCTCATATGCAGGCAGCACAGTAATTTTACAAGGTTGTAATCCTTTGGTTCCGGTT
GGTACAGTGTTGGAAATTCAACGTTTGCTTCCTGTTTCGACTGATCCTGCTCAGTTGCACTCGGTCGAAGA
CTTGCAGTTACCAGAGGCAGTGATGCGACTGCTTCGCAACTTTCAGGATGTGTTCCAGCCACCATCTGCC
CTGCCCCCGACTCGATCGTGTGATCATGCTATCCCTTTGATTGAAGGAGCGCGACCAGTCAACATCCGAC
CCTATCGATTTTCACCAGCAATGAAAGATGAGATGGAGTCTCAGGTGGCTGAAATGTTGCAGAGTGGTTT
GATTCAGCACAGTGATAGTGCATTCTCATCCCCAGTGTTGCTGGTCAAAAAGAAGGATAATTCATGGCGT
TTTTGCGTGGATTACAGACATTTAAACGCCCTTACAGTTAAATCCAAGTACCCTGTTCCCATCATCGATGA
GTTACTTGACGAGCTGTTTGGTGCCTCCTGGTTCTCCATATTAGATCTAAGAGCGGGCTTCCACCAGATCT
TACTCAAAGCAGGAGAGTCACACAAGACTGCCTTCCAGACACACACAGGCCACTATGAGTTTCGAGTTAT
GGCTTTTGGTCTCACTGGAGCTCCTGGTACATTCCAGAAGGCTATGAACCACACCCTAGCTCCCCTGCTTC
GGAAATGTGCTCTTGTGTTTTTTGATGACATTCTGGTTTACAGTGCCTCCTTGGAAGAGCATATTCATCAC
TTACAACTGGTTTTCGAGTTGTTAGCCCAGGATCAATGGAAAGTTAAGCTCTCCAAATGTTCTTTTGCACA
ACAGCAAGTCTCCTATTTGGGCCATGTCATCAATGCTCAAGGTGTAGCTACAGATCCAGTTAAAATTCAA
GCTATTGCCGAGTGGCCAGTGCCAAACTGTCTCAAGGATTTACGCAGTTTTTTGGGCCTTGCCGGCTACT
ACAGGAAGTTTATTAGGCATTTTGGGGTCATTTGCCAGCCCTTGACCGCTTTACTGAAGAAGGGCAGTCT
GTTTGTTTGGACAAATGAGCATGGTATTGCATTTGACACTTTAAAAAAGGCTTTGGTGACAGCTCCTGTTC
TTGCATTACCTGATTTCACTGCTGTTTTTCTCGTGGAAACTGATGCTAGTGCTGTCGGTGTGGGTGCAGTC
CTCATGCAAAAAGGTCATCCCCTCGCCTTCCTTAGTAAGGCCCTGGGTCCCAAGTCTAGGGGATTGTCAA
CATACGAAAAAGAGTACTTAGCAGTGATCTTGGCTGTTCAACAGTGGCGCCCCTACCTGCAACACCAGGA
GTTTGTTATCCTCACAGACCATAAAAGCCTCACACAGCTTAATGAGCAGCGCCTCCATACTCCTTGGCAAC

ACAAGGTATTTACCAAGCTCCTTGGTTTGCAGTACAGGGTCCAATACCGTCCGGGCACTGAGAATCGTGT
TGCTGATGCGTTGTCGCGGTGTGCCACCGCTGATCTGCACGCTCTGTCATCAGTCGTGCCTCAGTGGTTA
CTGGACGTGCAGGCCAGTTATACACAGGATGCTGCAGCTCAATCCATTCTGGCCAAACTTGCAGTCGACC
CAGCGGCTCTTCCAAACTATGTACTGAAGGATGGACTGATTCGGTATAAAGGTCGTATATGGTTGGGTGC
CGATGAACAACTACACCGCCGAGTTGTTTCTGCGTTACATTGCTCAGCCGTCGGTGGCCACTCGGGCTTC
CCTGTCACTTATAGCCGAGTCAAGCATCTGTTCGCTTGGACCAAGCTCAAGTCATTTGTTCAACGCTTTGT
CCAGCAGTGTCAGATCTGTTTACAAGCTAAACCGGACAGATCAACTCAGGATGCAGCGTCAGGCCGACA
AAAACAGGTCTGAGCGCTCGTTTGCCATCGGTGATTGGGTGTACATGAAACTCCAGCCCTATGTGCAGTC
ATCTGTCTCGCTGCGAGCTCATCATAAGCTGAGTTTTAAGTATTTTGGCCCATACAAAGTCACAGCTCGGG
TTGGTAAGGTTGCATATCGTTTGGACCTGCCTGTCTCCAGTAAGATCCACCCGGTGGTCCACGTATCACA
ACTGAAACTGGCAGCAGGTTACAAAGGCCCTGCTCCAGCAACTCTGCCTCTGGATTTACCTGAATTCAGT
GTTCCTGTTCAAATCCTTCAGACGCGAGGACTTACCAAAGGCAACCGTCTCGTACAGCAAGGCCTTGTGC
AGTGGTCCGGCTTACCTTCCGACCTGGCCACCTGGGAAGATCTTGAAGCTTTGCGTCAATGTTTTCCTGGT
GCTCCTGCTTGGGGTCAAGCAGGATTTCAAGGGAGCGGGAATGTTACCAGCCAGGCGTCAGTATCTGAT
CACCCACAGAAGCAGCCCCAGCGTCGCAAGCCTGGAACGCGTGAGCGTCGTGCCAATGTCCGCGTCACA
GGACCGGAGTGGGTGTAGTTGTGTGCGTGGTGTTCCTGTCCCTGTTTGATGGACCAGCTGCGTCTGGTG
GAGGGTTTCTTTATGTATCGCTAACTCTGAACTGGGAAGGCAGGAAACAAACAGAGAGAAAATATCTAT
TCCGGCCAGTTCCTATCTCTCTCTCTGTCGTCCTCCTCCTCCAACCTTTTCCCCATTTCCTTCTGTTAGTATAG
ATCAGAGCTGCTAACACCTCC

Tzi8

>chr5:122322345-122326261

TGGTTCGATTGCTCCATGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTTGCGTGGAGT
TGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATAGTGGGCCGG
CCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAGAAACAGAAT
CCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGGCGCGGCTCA
CGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCATCTGGTATCGGATTCGGGGTCGGCTGCGT
CTTCATCCTCTTCCACGTGCTGTTCCTCGTCTTTGGCGGTTCGTGGGCGACTTCTTCTTCTGCTGGGTCCAT
CCAGTTCATCACTACGGCGCCCGCAACCAAACCATCGGTACAGACTCGTTTGGTGTTACAGGCTATGGCG
GATGGGGAGCGAGTTAGTCAGGAGAATTGGAGTCGTGTTATTGAGACGTTGGAGGGAGTGAGTAGCAG
GTTGAAGGATGTGGAGAGGGTGCAGCAACAAATCATGATGCGGACTGAATTAGCGGCATCAGTTGCTG
AACAGGCTGCGAAGGAGAGAAATGATTTGGGTCGTAAATTGGAGGAGACTGGAAGGGGAATGATATTT
GGGCTTTAGTGGTGTTGGAAGAGGTGTCTGCAGATCAGGGAGTACAGCAGCAGGGGGATCTTCAGCAG
TTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTACCTCCTGCTAGGGTGTATGATCATCA
TATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTATAGGTACTCTCCTCTTCACAAGGATG
AAATTGAGAGGCAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATTGTTCCTAGTGTAAGCCCGTTTGCTTCT
CCGGTATTGTTGGTTAAGAAGAAAGATAGCACTTGGAGGTTTTGTGTGGACTACAGGAAATTAAATTCTT
TGACCATTAAGAACAGATTTCCTATGCCGCTTGTAGAAGAGATTCTGGATGAACTGGCTGGTACTAAATA
TTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAATGGGGGAAGGTGATGAGCATAAAACA
GCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCCTTTTGGGTTAACCAATGCTCCTGCAAC

CTTTCAATGTGCTATGAATTCTGTTTTGGCTCCTTTTCTTCGGAAATTTGTCATGGTTTTCATTGATGACATT
TTGGTCTATAGTGCTACATGGGAGGAACACTTGGCTCATATTAAGATGGTGCTGCAGACTTTGAGAGAG
CACCAGTTCTATCTGAAGAGGAGTAAATGTGTGTTTGGGAGGAATGAGCTCATTTATTTGGGACATATTA
TCTCTCACACAGGAGTGGCCACTGATCCTGCTAAGACAACAGCCATGGAGAAATGGCCTGTTCCTCAGTC
AGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTATAGGAGATTTGTGAGACATTATGGGAT
TATTGCAAAACCTCTTACTAATCTGTTGAAGCATAAGGCTTTTCAGTGGAGTCCTGAGGCTGATCAGGCTT
TTCAAGCGCTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCCCTACCCAGGTTTGATGAGTTATTTGTT
GTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGATGCAATGTGATAGACCAGTGGCATAT
CTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTATCCATTTATGAAAAAGAGTTTTTGGCTCTAATCA
TTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATTTGTGATTAAAACTGATCATCAGGCTT
TGAGTTTTCTGGATGATCAGCTTTTGCAATCTGATATGCAGAGGAAGGCTATGGCTAAGTTGATGGGGCT
TCATTTCAAAATTGTTTACAAAAAAGGCAAGGACAATGTGGCTGCCGATGCTCTATCCAGGATTGGTCAT
GTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGGGTGCAAGAACTCCTGAATGCTTATGTG
ACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGGCCAGTCCTAATGAGCAAGGTTTTTCTT
TACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGCTAACAATTCTGCCCTTCAAACTAAGTT
GATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGTGGTGAGGCCACTTACCAAAGGATCAA
GAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAGTTTGTGCAGCAATGTGAGGTGTGTCA
GCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAACCTTTACCAGTGCCTCAAGGGGCTTGG
CAGGATCTTACATTGGATTTCATTGAAGGTCTACCTAGATCTGAGGGATATGATACTATTTTGGTCGTGGT
GGATAGATATTCCAAATATGCTCATTTTTTTCCCCTCAAGCATCCTTTCACTGCTGTTGGAGTGGCACAGG
TGTTTCTTGACAATATTGTCAAGTTGCATGGGACACCAAAGTCTTTGGTTTCCGATAGAGACGAAGTGTTC
ACTAGTTTGTTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCAAATTGGCTCTCACTACTGCATATCACCCT
CAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGAAATGTATCTTCGCTGTTCTATTCATGAGG
CACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAATTTTGGTACAATACTTCATATCACACAGCTTTG
GGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTGTTTGCTGCTGCTCCTATGCTTTCTGGGGT
TGCTCCTAGTTCTGTGACTGAGGTGTTGGCTGATAGACAGTTACATGCTAGTATGCTGAAGGAGAAGTTG
GCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCATAGATCTGACAGAGAGTTTCAGATTGGA
GATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCGGTGGTTAGTCGACCCTGTCCTAAATTGG
CCTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGGGATGGTTGCTTATAAGTTGGATTTACCA
GTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGTTGAAGGCTTTTACTCCCAATTATTCTCCGGTG
TTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGATTATGCCATTGGCTGTGGTGGAAAGGAG
GATGGTGAAGAAGGGTAATCATCCTGTGGTGCAAGTAAGGGTGACTTGGTCCAATATGCCAGCTGAGGT
TACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTTCCTACAGCTCTTGATTGGGGACAATCAAGC
TCTTCCGGGGGGGGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGT
TTTGCGTGGAGTTGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAA
TAGTGGGCCGGCCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAAT
AGAAACAGAATCCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTT
GGCGCGGCTCACGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCACTCCAGTGATGATGCATG
GT

Ep1

>chr7:151979168-151983106

NNNNNTTTGTAATACTCAAAATTGTATAAAAGGATTATATAGTGAATTCCTTATCTTATGTGCCATATTTG
CATCATAAAAGAGCATACATGAAAGTTGGAGATGAATTAAAACAATGATAAAAAGAATATGCATCATG

TTGGAGTTTTATATGTTTGCACATTAAATAAAATAATAAGGATAATAACAATGAGATAATAATTCTAGAG
GTTGGATCAAACCCTAAATGGAAATTAGAGTTTGGAAATAAGAAAAAGAGAAGGGATAAAAAAANNNN
NNNNNNTAGCGGCATTCTTAGTAGATGGGAAGTTTGTTGGTGTCAACATCCGTGTGGTTAGAGATTTTC
CGGATGTCTTCCCAGAAGAGTTACCAGGGATGCCACCATATAGGAAAGTTGAGTTTGTCATAGATCTCTT
ACCTAGGACTACTCTTACTTCCAAACGGCCATACGAGATGTCCGTAGAAGAACTAAAGGAACTTAATAAG
CAATTAACAGAGTTACAAGAGGCTAGGTACATTCATTCGAGTTCCTCACCTTGGGGAGCACCGGTGTTAT
TCGTACAGAAGAAGGATGGATCACAACGGATGTGTGTGGACTATAGATCCCTTAATGATGTTACAGTGA
AGAACAAGTATCCGTTACCCCGTATTGAGGATTTGTTTGATCAGATAAGAGAAGCTAGAGTATTCTCGAA
GATTGATCTTCGATCGTGTTACCATCAGATGAGGATTAGACCATCGGATATTCCCAAGACGGCTTTCTCG
ACTCGATATGGATTATATGAGTTCTCTGTTGTGCCATTTGGATTAACCAATGCACCAGCTTATTTTATGGA
TTTGATGAATAAGGTGTTTATGGAGTATTTGGACAGATTCGTCGTAGTGTTCATCGACGATATTCTTTTTT
ATTCCAAGAGTGAAAGTGATCATGAGGAACATCTGAGATTGGTGCTACAGAAGCTACGAGATAATCAAC
TCTACGCCAAGTTTAGCAAGTGCGAGTTTTGGATTGACAATGTGTCATCCCTTGGACATGTTATTTCTAAT
GGAGGAATAGCAGTGGATCATCCTAAAGTAAAGGAGATAATGGAATGGACCGTGCCCACTACCGTCACT
GAGATTCGGAGTTTCTTGGGACTTGCAGGATATTATCGGAGATTTATGGAAGGATTTTCTAAGATTACTA
AGCCTATCACCTCGCTTCTGAAGAAAGGAAAAGAATTTAAGTGGGATGAGAAATGTCATGAGAGCTTTG
ATCAATTGAAGGAGAGACTAATGTCGCCACCAGTATTGATTATGCCAGATCTGCAGAAGGGGTTTGACAT
TTATTGTGATGCATGTGGCCAAGGACTTGGATGTGTGCTTATGCAAGAAGGACATGTGATCGCCTATGCA
TCTCGTCAGTTGCGGAAACATGAATTGAACTACCCCACTCATGATTTAGAGCTGGCAGCAATTGTGCATG
CACTGAAGATTTGGAGACACTACATTATGGGAACTAAGTGCCAAGTATACACGGATCATAAGAGTTTGA
AGTACATACAGAAAGATCTCAACCTTAGGCAACGCTGTTGGTTGGAGCTTATTAAGGATTACGATTTGGA
GATTCACTATCACCCGGGCAAGGCAAATTTGGTTGCAGATGCCTTGAGTCGGAAGGAGCATGTTCACTCC
GCTTTCGTTGTCCAGCTACCTGATGAATTAGCAAAAGATTTTGAGAGACTCAATTTGGGAATTGTTACTCA
TATTGAAGGAGTCACCATCGAGTTGGAACCCACCTTGGAGCAAGAGATTCGTAAAGGTCAGGTTGGTGA
TGCTAAGATTCAAGAGATCAAGGATCTTATAACTGAAGGAAGAGTTCCGGAATTCACGGAGGATGAGCA
AGGTACTATATGGTTCAAAGATCGGATATGTGTTCCAGAGATTGGTAGCCTCCGTGAGACTATATTGAAG
GAAGCCCATGATTCTGATTATTCTATCCACCCTGGTAGCACTAAGATGTATCAGGATTTGAAGCAAAAGT
ATTGGTGGTATGGACTAAAAAGAGATGTTGCTGCTCATGTGGCTAAGTGTGATGTGTGTCAAAGAGTTA
AGGCTGAACACCAGAGGCCAGCTGGACTACTGCACCCACTGAAGATACCCGAGTGGAAATGGGAAGAA
ATTGGTATGGATTTTATTGTTGGACTACCCCGCACTCCAGCAGGATATGACTCTATTTGGGTAATTGTGGA
CAGATTGACAAAGGTGGCTCATTTTATACCTGTGAGGACTAATTATACAGGAGCCAANNNNNNNNNNG
GAGGAGCCACAAGGCGAGGAGTTTGATCTGATCTAGGTGGCGTTTCTCAGTTGACATTGGCGCCGACGA
TCCTTAGTTCGTTTTATGTTTATTCTTTTATTTTGTAATAAGTCTTCCGCTATGTAATAAATACTCTGATGTTT
GATGACATTTATCTCTATACACTCTGTTATTGTATATGTTGTCTTCTTGGCGCATGTATGAGATGCACCCGG
CTTTGTTCCTTAAAACCGGGTGTGACAGNNNNNNNNNNTGAAGATACCCGAGTGGAAATGGGAAGAA
ATTGGTATGGATTTTATTGTTGGACTACCCCGCACTCCAGCAGGATATGACTCTATTTGGGTAATTGTGGA
CAGATTGACAAAGGTGGCTCATTTTATACCTGTGAGGACTAATTATACAGGAGCCAAACTAGCAGAATTG
TATATGACTCGAATAGTTTGTCTACATGGAGTGCCTAAGAAGATCGTGTCTGATCGAGGATCACAGTTTA
CTTCTCGATTTTGGAAGAAGTTGCATGAGAGCTTGGATACAAAGTTGAATTTTAGTTCCGCCTACCATCCC
TAGACTGATGGACAGACAGAGAGGACTAACCAAGTGCTGGAAGATATGTTAAGAGCTTGTGCTCTCAAG
CATGGTGGCAGTTGGGAGAAGAGTTTACCATATGCTGAGTTTTCTTATAATAATAGCTATCAGGCTAGTT
TGAAGATGTCACCATTTGAAGCTCTATATGGTAGAAAATGCAGGACTCCACTGTATTGGGATCAGACTGG
AGAAAGACAGTTCTTTGGGCCTGAACTTATACAGGAGGCAGAAGAACAAGTGCGAATGATAAGAGAAA
ATTTGAGAATTGCACAATCAAGGCAAAAGAGCTATGCTGATAATAGAAGGAGACCACTGGAGTTTAAGG

AGGGTGATCATGTCTATTTGAAGGTGTCACCACTCCGAGGTATGAGGAGATTTAAAGTTAAAGGAAAGT
TGTCCCCTCGCTTCATTGGACCTTTCTTGATCTTAAAGCGAGTGGGAGAAGTTGCATATCAACTGGAGTTA
CCCAATCATCTGTCAGATGTACATGATGTGTTCCATGTGTCCCAGCTAAAGAAGTGCCTCAGGGTACCTG
AGGAGCAGTTACCAATGGAAGACCTTAGTGTTCAAGATGACCTGACTTATGTTGAGTACCCAATCAAGAT
TCTGGACACTTTGACTCGAGTTACAAGAAATAAGGTGATCAAGATGTGTAAAGTGCAATGGAGTCACCA
CGGTGAAGATGAAGCAACTTGGGAAAGAGAAGAAGAGCTTCGCATAGACTTCCCCCATCTTTTCCCTAG
ATCTTCTTAAATCTCGAGGACGAGATTATTTTTAAGGGGGGTAGGATTTGTAATACTCAAAATTGTATAAA
AGGATTATATAGTGAATTCCTTATCTTATGTGCCATATTTGCATCATAAAAGAGCATACATGAAAGTTGGA
GATGAATTAAAACAAATGATAAAAAGAATATGCATCATGTTGGAGTTTTATATGTTTGCACATTAAATAA
AATAATAAGGATAATAACAATGAGATAATAATTCTAGAGGTTGGATCAAACCCTAAATGGAAATTAGAGT
TTGGAAATAAGAAAAAGAGAAGGGATAAAAAAATATAT

>chr4:8015632-8020208

NNNNNTGTCAGGACCGGGTGAATCGGTGAAGTATTAGTGCTCTAATGGGCCTGTTAATCTTGGCCTTGT
AATCCTAGTGTTTGGGCCGGTGTTGACAATCATAGATTGAGTTGGTGGCACTGTTATAAGCCACACCGTA
GCACTTGCTGGTCAGGAAATGAGAACAGAAAAATCTACACATACCTGTCTTCTCCTTCCATTTCCTTCCGC
TTCTATTTCGTTTCTTCTCTTCTCCAGTTTCCTCTCCGGTGATCACCGGCATCTCGCCNNNNNNNNNNCAT
CTGAAAGGGAAGGCTGCACACTGGTTTCAGGCATATAGAGATTCATTGAACTTGATGGATTGGGCTGAG
TTCAGAGTGGCTGTGTTGGAAGAGTTTGAGGTCTCTACTCATTCAGATAAGATGCAGGAGTTGCTAACAT
TGAGACAAACTGGTTCGATGTTGGAGTATAAAACCCATTTTGAGAAATTGGTGTACCATATAAAGTTGTT
TGACAAGGCCATCAGTGACACCTTACTGGTTACACAATTTGTTCTGGGTCTCAAATGGGAATTGAGAGGT
GGGGTGGAAGTACAATTTCCCAAGACAGTGTCAATGGCTGCTCAATTAGCATTCAAGCATGAAGCATTGC
AGGCAAAACAGACATCAAGGAGACAGGCAGTGGGGAAGCAGGGTGCAGCAATACACAATAAAGATGG
AGTAGGAGTCACTGGTGAATTGTGGAAAGCTAGACAACTGAAAGAGTATAGGAGAATTAATGGGTTGT
GTTTTAAGTGTGGAGAGAAGTTCATCCCAGGTCACAAATGTAAGGTGGTTGTGCAACCTCAGCTGAATGT
CATGGTTGCAGAAGGTCAGGATGGAGGTCAGATCCTGCAAGATGAGGTGCTAGAATTATTGGAAAGGG
TTGAGCACCGGAATGAGGAGGAAATGACCTTGTCTCTGAATGCCATGAATGGTACTGACACTTCAAAGT
GTATCAGGCTTAGGGCAATGGTTAAGAACCATACCATTTTGCAGTTGCTTGATTCAGGGAGCTCCAATAC
TTTTATCAGTGAGATAGCTCTAGCTAAAATGGGTTGTAAGGTCCAGGAAATACAACCATTAGATGTGAAG
GTAGCCAATGGACAAGCTATCAGGTGTGAAAAGAGAGCTTTGCAACTGGAATGGTGCTGTTGTGGAAAG
ACTTTTTGTGCTGATGCTTATGTGCTCCCTGTCACTGCTTTCGACATGATTTTGGGTATGGATTGGCTAGA
GCAGCACAGTCCAATGTTATGTGCTTGGGACAAAAAGTGGGTGGAATTTGAGGAAGACGGTGTGCCCAT
CAGGTTGCAAGGGGCTTTGGATAGACCAGAGAAGGTTATTCGTGAGTCCAGTGGTGATCAAGTGTATAA
ATGGGAAAAAGGAAATGAGGTGTGGGCTGTGGCATTGATACAGAAGGTGGAGGAAGGGTTACAGCCA
AAGGTGGAGGTAGTTCCAGCTGAAGTTCAAGAGGTGTTAACCCAGTTTCATGATGTGTTCAAAACTCCAG
ATGTTTTACCGCCATCAAGAGATTATGATCATTCCATCACATTGTTACCTGGAACAGCTCCTGTGAACTGC
AGACCATATAGATACTCACCTTTACAGAAAGATGAAATTGAAAGGCAAGTGATGGAGATGCTCAAGTCA
GGACTCATCACTAGGAGTGTCAGTCCCTTTGCTGCACCTGTATTGTTAGTAAAGAAGAAAGATGGGAGTT
GGCGTTTTTGTGTTGATTATCGAAAGCTGAATGAAATCACTGTCAAGAATAAATTTCCCATGCCTGTGATA
GATGAGTTCTTGGATGAATTGGCAGGAGCCACATATTTCTCCAAGCTGGATATGGCATCAGGGTTCCATC
AGATAAGAATGGCTCAAGTTGATGAGGAAAAGACAGCCTTCAAGACGCACCATGGTCATTTCCAATTCA
GAGTTATGCCATTTGGGCTTACAAATGCCCCGGCCACATTTCAGTGTTTGATGAATGCAGTATTCCAGCAC
TGTATGAGGAAGTTTGTGTTGATCTTCATGGATGACATCTTGGTTTACAGTCCCACTTTGCAAGCTCATGT
GGTCCATTTGCAAGAAGTGTTCTCAGTGTTGCAACAACACAAGTTGTATGCGAAGAGAAGTAAGTGTTCT

TTTGCTGTGACACATATTGAATATTTGGGGCATGTCATATCTGCTAATGGGGTTGCCACTGACCCAAGCA
AAACTGAAGCAATGCTCAAATGGCCAGTACCCACTTCTCACACAGAGTTGAGAGGATTCTTGGGACTCAC
TGGGTATTATCGGAAATTTGTTCAGAATTATGGTATAATTGCCAAACCTCTGACAACTATTTTGCAGCTGA
AATTATTTGAGTGGCCAGAGTCTGCTCAAATTGCTTTCGATGAACTCAAGAAGGCTATGGCATCCACTCCT
GTTCTAGCTTTGCCAAATTATGAAGAAACTTTTGTGGTGGAAACAGATGCATGTGACAAGGGCATTGGTG
CTGTACTATCGCAAAAGGGACACCCGGTGGCTTTCTATAGCAAGGCATTAGGGGTTTCCAATTGCAAGTT
GTCAATCTATGAAAAGGAGTTCCTGGCCATTCTTATGGCAGTGGATAGATGGAGATGTTACCTACAGAG
GAAACCTTTTGTAATTAAAACAGACCACAAAAGTCTGACTCACTTGCAGGACCAAACTCTGGTCACTGAA
TTACAAAGGAAGGCAATGGCCAAACTGTCTGGACTTCAGTTTCAGATCCAGTACATGAAAGGGGTCGAG
AACAAGGCTGCTGATGCACTATCGAGGGTGGCACATCAGTTGGAGTTTGCTGCTTTGTCTGTGTGTACTC
CAATGTGGGTTCAAGAAGTTCTCAAATCCTATGAGTTGGATTGTAGTGTGCAGAAATTGCTGGCTGAGTT
GGCTGTGCACAGCCCCAATGAGCAAGGGTTCAGGATTCAAGAAGGTCTGGTATATAAGCAGAACCAATT
GCTGGTGGGTCAGAATTTGGGGCTACAGACAAAGATTATCAAGGCTCTTCATGCATCTGCAATAGGAGG
ACATTCTGGAGTACAAGCTACTTATCAGAGGGTCAAGAAATCCTTTTACTGGCCTGGGTTGAAAAGGGCA
GTGGAGTTATTTGTCAAGCAGTGTCAAGTGTGCCAACAAGCAAAACATGAGCATGTCAAGTCACCTGGG
TTACTTCAGCCTCTACCCATTCCACAAAATAGTTGGCATGACTTGTCTATGGATTTTATAGAGAGTCTACCT
AGATCAAATGGATTTTCAGTGATCCTGGTGGTGGTAGACAGGTTAACAAAGTATGCACATTTTGTTCCAG
TGAAACACCCTTACACTGCGCAACAAATAGCCCACATTTTTTTCTCCAATGTGGTTAAACTGCATGGGTTG
CCAAAATCCATTGTGTCAGACAGAGACAGAGTTTTCACTAGCCACTTTTGGCAAGAATTGTTCAAGTTGG
TGGGCACTCAGTTGCATCTTAGCTCGGCATACCATCCACAATCTGATGGACAGACGGAGAGAGTTAATCA
ATGTTTAGAAATGTTTTTGAGGTGTGCGGTGAATGACACTCCCTCCCAATGGGTGAAATGGTTAGATTCG
GCTGAACTATGGTACAATAGCTGCTACCATTCATCTCTAAAATGTTCTCCATTCAAGGCTTTGTATGGGGT
AGAGCCAAACTTGGGATTTTTGCCTATCTCACTGTCTGATGGTTCAACTGAAGCTGCAGTATCCTTGCAAC
AGAGACATGACTATCTTGACATGATTAAAACTCACTTGGCCTCAGCTCAAAACAAGATGAAGTTTTATGC
AGACCAGCAAAGGACATTTAGACAGTTTGCTGTGGGGGATAAGGTGTTTGTCAAGCTTCAGCCGTTTGC
CCAAACATCGGTGGCTTACCGACCTTGTGCTAAACTCTCTTTTAAGTTCTTTGGACCATTTCAGGTGATAG
AAAAGATTGGTTCGGTGGCTTATCGTTTGCAGATGCCGGACACAGTGGCAATACATCCTGTTTTCCATGT
GTCGCAACTTAAGGAGTTTGTGCCAGATCATACTCCGGTGTTTACTCAGCTTCCTCCGGCAGCCTTTTCAG
ACAACTCTCCACCGGTTCCTGAAGCCATTTTGGATCGTCGTTTGGTCAGAAAGGGGTCTGCTGCAGTAGT
CCAGATTCTTATCAAATGGAGAGGATTGCCTTCCTCCATGTCATCATGGGAAGATTATACCGTCCTGAAG
GAGAGATTTCCTGCTGCTGCAATCTGGGGACCAGATGCATCTCAAGGGGAGGATGATGTCAGGACCGG
GTGAATCGGTGAAGTATTAGTGCTCTAATGGGCCTGTTAATCTTGGCCTTGTAATCCTAGTGTTTGGGCC
GGTGTTGACAATCATAGATTGAGTTGGTGGCACTGTTATAAGCCACACCGTAGCACTTGCTGGTCAGGAA
ATGAGAACAGAAAAATCTACACATACCTGTCTTCTCCTTCCATTTCCTTCCGCTTCTATTTCGTTTCTTCTCT
TCTCCAGTTTCCTCTCCGGTGATCACCGGCATCTCGCCACGGG

>chr4:258921835-258925715

NNNNNTGTAATATCCAAATTATAAGACTTGTAAAAGAATTGGTCTCTGTATTTGTATTGCCATCTACCTCT
TGAGACATGTTACTAAACATAAGTAAAAGAGCATAATTCCTTATTGAAACATTAAGAGAAAGCCATTTCA
TCATGTTGGAGTTTGGTTGTGCATTAAATAACAGTGAAAATAAAACCAATATAAATAAAAAAATAACAAG
TTGAAGATTTGGTCTGAAACCTAAATTTAAAAATGGAAAAAAATGAGAAAAAGGAGAGAAAAGAAAAG
GGAAGGGGTATAAAATACAAATAGAAATGAAAATATAAAAGAAAAAGGAGGACTCATGCGCATGGGCC
GATTGTCCCTCATTTGGCCCATTTGCCATNNNNNNNNNNTATAGTGTTAGAAGAGAGCACTATAGACGT

GATCCTTGGTATGTCATGGTTAAGAAAGGCAAAGGCAGTTTGTACACTGTGCTAAAGGAACCATATAACT
CACCAGTTCCAAAGGAGAAAGATTTGAAGTTGGAATTGCAGTAATTACCGCCACCAGACTAGCGGCATTC
TTAGTAGATGGGAAGTTTGTTGGTGACAACATCCGTGTGGTTAAGGATTTTTCGGATGTCATTCTAGAGG
AGTTACCAGGGATGCCACCAGATAGGGAAGTTGAGTTAGTCATTGATCTCTTACCTGGAACTGCTCCTAT
TTCTAAACGACCATATAGGATGTCTGTAGAAGAGTTAAAAGAACTCAAGAAGCACTTAACGGAGTTGCA
AGAGGCTGGGTACATTCGTCCGAGTCTTCACCTTGGGAGAGCGCCAGTTTTGTTTGTACAGAAGAAAGA
TGGATTGCAAGGGATGTGCGTGGATTAAAGGTCCCTTAATGGTGTCACTATGAAGAACAAGTATACATTA
CCCCGCATTGAAGATTTATTTGATCAGATGAGAGGTGCAAGGGTATTCTCGAAGATTGATCTCCGATCAG
GTTATCATCAAATGAAGATTTGGCCATCGGATATCCCCAAGACAGCTTTCTCGACTCGATATGGATTATAT
GAGTTCTCTGTTATGTCGTTTGGATTAGCTAATGCACCAGCTTATTTTATGGATTTGATGAATAAGGTGTT
TATGGAGTATTTGGACAGATTCGTCGTAGTGTTCATCGACGATATTCTTTTTTATTCCAAGAGTGAAAGTG
ATCGAGGAACATCTGAGATTGGTGCTACAGAAGTTACGAGATAATCAACTCTACGCTAAGTTCAGTAAGT
GCGAGTTTTGGATTGACAAGGTGTCATTCCTTGGTCATATTATTTCTAATGGAGGAATAGCAGTGGATCC
TGCTAAAGTGAAGGAGATAATGGAGTGGAGAGTGCCCACTACAGTTACTGAGATTCGGAGTTTCCTGGG
ACTCGCAGGATATTATCGGAGATTCATTGAAGGGTTCTCTAAAATTGCCAAGCCTATGACATCGCTTTTG
GAAAAAGGAAAAGAATTTAAGTGGGATGAGAAATGCCAAGAAAGCTTTGATCAATTGAAGGAGAGATT
GATGTCACCTCCGGTGTTGGTAATGTCAGACCTACAGAAAGGATTTGATATCTATTGTGATGCATGTGGC
CAAGGATTGGGATGTGTGCTCATGCAAGAAGGACATGTGATCGCCTATGCGTCTCGTCAATTGCGGAAA
CACGAGTTGAACTACCCCACTCATGACTTGGAACTGACAGTCGTTGTGCACGCACTTAAGATTTGGAGAC
ATTACATTATGGGAACTAAGTGCCAAGTATACACGGATCATAAGAGTTTGAAGTATATATTCACTCAGAA
GGATCTCAACCTTAGGCAACGCCGTTGGTTGGAGCTTATTAAAGATTACGATTTGGAAATTCACTATCACC
CGGGCAAGGCGAATTTGGTTGTAGATGCCTTGAGTCGAAAGGAGCACGTTCATTCAGCTGTTGTTGCCC
AGCTACCCGATGAGATTGATGAGGATTTTAGGAGACTTAACCTGGGGATAGTTGATCATGCTGAAGGAG
TTATTATTGATGCGGAACCTACCTTGGAGCAAGAAATCCGCAGAAGGATAGATTGGTGATGCTAAAATAC
AAGAGATTAAGGATCTGATTACTGAAGGTTGAGGTTCGGAATTCACGGAAGATGAGCAAGGCACCGTAT
GGTTCAAGGACAAGATATGTGTTCCTGATATTGAAAGCCTTCGTGAGACTATTTTAAAGGGAGCCTATGA
CTCGGATTATTCTATTCATCCTGGTAGTACCAAGATGTATCAGGATTTGAAGCAGAAATACTGGTGATAT
GGACTGGAGAGAGATATGGCTACACATGTGGTTATGTGCGATGTGTGCCAAAGAGTTAAGACCGATCAC
CAAAGACCAGCTGGACTATTGCACCCGTTGAAGATACCCGAGTGGAAATGGGAAGAAATTGGTATGGAT
TTTATTGTTGGACTACCCCGCACCCCTGCTGGGTATGACTCTATTTGGGTAATTGTGGATAGATTGACAAA
GGTGGCTCATTTCATACCAGTGAGGACCAATTATACGGGAGCCAAGCTAGCAGAATTGTATATGACTCG
GATAGTTTGTCTACATGGAGTACCTAAGAAGATCGTGTCTGATCGAGGATCACAGTTTACTTCTAGATTTT
GGAAAAGTTTTCATGAGAATATGGATACGAAGTTGAATTTTAGTTCGGCCTACCATCCACAGACTGATGG
GCAGACTGAGAGGACTAACCAAGTGCTGGAAGATATGTTAAGAGCTTGTGCCCTTAAACATGGTGGAAG
TTGGGATAAAAGTTTACCTTATGCTGAGTTCTCCTATAATAATAGCTATCAGGCCAGTTTGAAGATGTCAC
CGTTTGAGGCTTTGTATGACAGAAAATGCAGGAGTCCTCTATATTGGGATAAGACTAGAAGAAGACAAT
CATTTGGACCTGAACCTATTCAAGAGGCAGAAGAACAAGTCCGTATCATTCGAGAGAATTTGAGGGTAG
CTCAGACCAGACCAAAGAGCTATGCTGATAATAGAAGAAGACCACTTGAGTTTGAGGAAAGAAATCATG
TGTACCTCAAGGTGTCACCACTTCGTGGAATGAGGAGATTCAAAGTCAAGGGCAAATTGTCCCCTCGCTT
TATTGGAACATTCTGAATTTTTAGGCAAGTTGGAGAGATGGCCTATCAACTCGAGCTACCTGATAATCTAT
CTTATGTGCATTCCATGTGTCTCAACTTAAGAAGTATCTCCGCGTCCCTGAGAAACAGTTACCAATGGAAG
AGCTCAGAGTTCAGGGAAATTTGGCTTATACGGAGTATCCTATCAAGATTCTTGATACTTTGACTCGAGTT
ACAAGAAATAAAGTTATTAAGATGTGCAAAGTGCAATGGAGTCACCACGGAGAAGATGAAGCAACGTG
GGAGAGAGAAGAAGAGCTTCGCATAGATTTTCCCCATCTTTTCCCTAGTCCTTCCTAAATCTCGAGGACG

AGATTATTTTTAAGGGGGGTAGGATTTGTAATATCCAAATTATAAGACTTGTAAAAGAATTGGTCTCTGT
ATTTGTATTGCCATCTACCTCTTGAGACATGTTACTAAACATAAGTAAAAGAGCATAATTCCTTATTGAAA
CATTAAGAGAAAGCCATTTCATCATGTTGGAGTTTGGTTGTGCATTAAATAACAGTGAAAATAAAACCAA
TATAAATAAAAAAATAACAAGTTGAAGATTTGGTCTGAAACCTAAATTTAAAAATGGAAAAAAATGAGA
AAAAGGAGAGAAAAGAAAAGGGAAGGGGTATAAAATACAAATAGAAATGAAAATATAAAAGAAAAAG
GAGGACTCATGCGCATGGGCCGATTGTCCCTCATTTGGCCCATTTGCCATCTCCG

>chr10:103596801-103601065

NNNNNTGTAATACTCAAAAGTGTAATAAAGAGAATAGAGATAATCTCAATAGTTGTGGCCTCTACTATTC
ATTACATGTGATCAACACTGGTTAAAAGTGGATAAATAGAATAAAGAACACATAATTAACCTGTGCATCA
TGCTGGATCTTACTTTGTGTGCATTTTTGTGATGATATAAGAAAATAGAAAAAAATAATAAGTATGTTAAT
ACTCCAACGGGAAATTAAATCCTAAAAGAAATTCTAGAGTTTGGAAAAGAAGAAAGGGAATACTATGTG
ACAAAAAAATAGAATAAAAATATATAATAAAATATATTATTTGTACATTAAGATTTGAATTTGTACATTTA
CTCAAAGGTACAACTCATAAAACTATCCAAATTAAAATTGGGAATTCAAAATAGAAGGAATAGAAAATAA
AATCAAAAAGAAAAAAAANNNNNNNNNNTTAAAATTATTTTCTCTATTAAAAATGTTGGTTGATTGTTCT
GATGATCAATGCTTGATTTGCTTCTTTGATTGATTGAAATAGTATAGACGGGCATCTTAGCATGTACCACC
ATAAGGTAATGAATTAGTTTTAGTAGGCGACACACTCAGCTAATAAATCCCCAAAGTAACACGCCTCGTA
ATTACATGCCTTGTCTACACTCCTTCCTTTCGTACCTTATGTTTCTAACCCAGATGGGCTACCGCTATAGTG
TTAGAAGAGAGCACTATAGACGTGATCCTTGGTATGTCATGGTTAAGAAAGGCAAAGGCGGTTTATTCA
CTGTGCTAAAGGAACCGTAGAACTCACCAATTCCAAAGGACAAAGATTTGAAGTTGAAGTTGCAGTAGC
TACCACCATCAAACTAGTGGCATTCTTAGTAGATGGGAAGTTTGTGGGTGACAACATCCATGTGGTTAGA
GATTTTTCCGGATGTCTTTCCAAAGGAGTTACCAGGGATGCCACCAGATAGGAAAGTTGAGTTTGTAATT
GATCTTCTACCTGGAACTGCTCCTATATCTAAACGGCCATACAAGATGTCTATAGAAGAGTTAAAGGAAC
TTAAGAAACAATTAACGGAGTTACAAGAGGCTGGGTATACTCGTTCGGGTTCCTCACCTTGGGGAGCACC
GGTCTTGTTTGTACAGAAGAAGGATGGATCGCAGCGGATGTGCGTGGATTATAGATCTCTTAATGATGTT
ACAGTGAAGAACAAGTATCCGTTGCCCCGCATTGAGGATTTATTTAATCGGATGAGAGGTGCTAGAGTAT
TCTCGAAGATTGATCTCCGATCCGGTTACCATCAGATGAGGATTAGACCATCGGATATTTCCAAGACGGC
TTTCTCGACTCGATATGGATTATATGAGTTTACTGTTATGTCGTTTGGATTAACCAATGCACCAACTTACTT
TATGAACTTGATGAATAAGGTGTTTATGGAGTATTTGGACAGATTCGTCGTGGTATTCATCGACGATATT
CTTATTTATTCTCAGAGTGATAGCGATCATGAGGAACATCTAAGGATGGTGCTACAGAAGCTACGAGATA
ACCAGCTATATGCCAAGTTCAGCAAGTGCGAGTTTTGGATTGACGAGGTGCCATTTCTTGGTCATATTATT
TCCAATGGAGGAATTTCAGTGGATCCTGCTAAGGTTAAGGAGATAATGGAGTGGAGAGTACCCACTACA
GTTACTGAGATTCGGAGTTTCTTGGGACTTGCTGGATATTATCGGAGATTTATTGAAGGATTTTCTAAGAT
TGCTAAGCCTATGACTTCACTTTTGGAGAAGGGAAAAGAATTTAAGTGGGACGAGAAGTGTCAAGACAG
CTTTGATCTATTGAAGAAGAGGTTGATGTCGCCACCGGTATTGGTTATGCCAGACCTACAGAAGGGGTTT
GACATTTATTGTGATGCATGTGGCCAAGGACTTGGATGCGTGCCTATGCAAGAAGGACATGTGATTGCCT
ATGCATCTCGTCAGTTGCGGAAGCATGAATTGAACTACCCCACCCATGACTTGGAACTAGCAGCAGTTGT
GCATGCAGTAAAGATTTGGAGACATTACATTATGGGGACCAAGTGTCAAGTATATACGGATCATAAGAG
TTTGAAGTATATATTCACTCAGAAGGATCTCAACCTTAGGCAACGCCGTTGGTTGGAGCTTATTAAGGATT
ATGATTTGGAGATTCACTATCACCCGGGCAAGGCGAATTTAGTTGCAGATGCCTTGAGTCGAAAGGAGC
ATGTTCATTCAGCTATTGTTGCCCAGCTACCCGATGAGATTGTTGAGGATTTCAGGAGACTTAACCTGGG
GATAGTGGCTCATACTGAAGGAGTTACTATTGATGTGGAACCTACTTTGGATCAAGAAATCCGCAAAGG
ACAAATTGGTGATGCTAAGATATAAGAGATTAAGGATCTGATTACTGAGGGTCGAGTTCCGGAATTTAC

GGAAGATGAGCAAGGCACCGTATGGTTCAAGGACAGGATGTGTGTTCCTGATATTAAAAGCCTTCGAGA
GACTATTTTAAAGGAGGCCCATGATTCAGATTATTCTATTCATCCTGGAAGTATTAAGATGTATCAAGATT
TGAAGCAAAAGTATTGGTGGTATGGACTAAAGAGAGATGTGGCTGCACATGTGGCTATGTGCGATGTGT
GTCAAAGAGTTAAGGCTAAACACGAGAGACCAGCTGGATTATTGCATCCACTGAAGATACCCGAGTGGA
AATGGGAAGAGATTGGTATGGATTTCATTACTGGATTGCCTCGCACCCAGAAAGGATATGATGCAATAT
GGGTGATTGTGGATAGATTGACTAAAGTGGCTCACTTTATTCCTGTGAAGACCACTTATAGAGGTTCTCA
GCTAGCAGAATTGTATATGGCTCGGATTGTGTGTTTACATGGAGTACCAAAGAAGATCGTGTCTGATCGA
GGTTCGCAGTTTACCTCAAGATTTTGGAGAAGCTTTCATGAGAACATGAGTACGAAGTTGAATTTTAGTA
CGGCTTATCACCCTCAGACTGATGGACAGACTGAAAGGACAAATCAAGTTTTGGAAGATATGTTGAGAG
CATGTGCCCTTCAGCATGGAGGAAGTTGGGATAAGAGTCTGCCTTATGCTGAGTTCTCGTATAATAATAG
TTACCAGGCCAGCCTGAAGATGTCACCTTTTGAAGCTTTATATGGGAGGAAGTGCAGGACTCCTTTGTAT
TGGGATCAGACTGGAGAAAGACAGTTCTTTGGGCCTGAACTGATTCAAGAAGCAGAAGAACAAGTCTGT
ATAATTCGAGAGAATTTGAGAGTGGCTCAAACCAGGCAAAAGAGCTACGCTGATAATAGAAGAAGACCA
CTGGAATTTGAAGAAGGAGATTATGTGTACCTCAAAGTGTCACCACTTCGTGGAATGAGGAGATTCAAA
GTTAAGGGCAAAATTGTCCCCTCGTTTTATTTGACCATTCAGAGTTTTTAGGCGAGTTGGAGAAATAGCCT
ATCAACTCGAACTACCGGATAGTCTATCGGATGTGCATAATGTGTTCCATGTATCTCAACTTAAGAAGTGT
CTCCATGTCCCCGAGGAACAGTTACCAATGGAAGAGCTTAGTGTTCAGGGTGATTTGACTTACACGGAGT
ACCCGATCAAGATTTTAGATACATTGACTCGAGTTACAAGAAATAAGGTGATAAAGATGTGTAAAGTACA
ATGGAGTCACCACGGTGAAGATGAAGCAACTTGGGAAAGAGAAGAAGAGCTTCGCGTAGATTTTCCCCA
CCTTTTCCCTCGTTCTTCTTAAATCTCGAGGACGAGATTATTTTTAAGGGGGGTAGGATTTGTAATACTCA
AAAGTGTAATAAAGAGAATAGAGATAATCTCAATAGTTGTGGCCTCTACTATTCATTACATGTGATCAAC
ACTGGTTAAAAGTGGATAAATAGAATAAAGAACACATAATTAACCTGTGCATCATGCTGGATCTTACTTT
GTGTGCATTTTTGTGATGATATAAGAAAATAGAAAAAAATAATAAGTATGTTAATACTCCAACGGGAAAT
TAAATCCTAAAAGAAATTCTAGAGTTTGGAAAAGAAGAAAGGGAATACTATGTGACAAAAAAATAGAAT
AAAAATATATAATAAAATATATTATTTGTACATTAAGATTTGAATTTGTACATTTACTCAAAGGTACAACTC
ATAAAACTATCCAAATTAAAATTGGGAATTCAAAATAGAAGGAATAGAAAATAAAATCAAAAAGAAAAA
AAAAGAGA

F7

NNNNNATATATATATATATATATATCATTTCAGGATAGTACTTTTATTTGAGCACTTGAACCCTGTGTGAA
ACCCATAAACAAATTTGAATTCAAATTATAGATTGGAGTGGGAGAATCAAAACAGGAATTAAAAAAATG
AGAAAAGGAATTCCTGGCGCCTTGGGCCGAATTCTTAGTCCCGCGGCCCATTTTCCTCCACCCAGCACAG
CCCAGCTCCTACTCTCCTTGCGTCACACTGACATGTGGGGCCAGCATGGTCAGCCACCCGCGCGCTATCCT
TCTCGCTGGCGGGTGGAGCCCACTTGTCGGCACTCTCTTCTCCGTCGTAACGCCCGCGCGGAATTGCCGC
TCGCGCTCGCACCTCACCCTGTCCCTGGGAGGCTGGGATGCGGACTCGCGTGCCAGCCGCCGATCGTGA
ACGTGCTATCGCTGCGCTGTGGGGCCCGTTCTCCTGTGCACGCCCCATTTCCGCAACAGCTCGCGCGGGA
GTCGCCGTCTTCGTCGGGGATCGTGGGCGACTTCTCTTCCCCCCAAATCCCGTGAGTCCGCGCCGCAATC
AACTCGCCTGGCTGTTATATGGGAGCTTAGGTCGCGATCTCCTCGGCCTTCCNNNNNNNNNNAGTTCCA
AAGGACAAAGATTTGAAGTTGAGTTGGTCATTGATCTCTTACCTGAGACTGCTCCTATTTCTAAACGGCTA
TACAGAATGTCCATAGAAGAACTAAAGGAACTTAAGAAACAATTAACGGAGTTACAAGAGGCTAGGTAT
ATTTGTTTGGATTCTTCACCTTGTGAAGCGCCAGTTCTATTTGTACAGAAGAAAGAAGAATCGCAAGGGA
TGTGCATGGATTATAGGTCTCTTAATGATGTTATTGTGAAGAACAAGTATCCGTTGCCCCGTATTGAGGA
TTTATTTGATCAGATGAGAGGTGCTAGAGTATTCTCGAAGATTGATCTCCGATCGGGATACCATCAGATG
AGGATTAGACCATCGGATATTCCTAAGACGGCTTTCTCGACTCGATATGGATTATATGAGTTCACTGTTAT

GTCATTTGGATTAACCAATGCACCAGCTTATTTATGAACTTGATGAATAAGGTGTTTATGGAGTATTTGG
ACAGATTCGTCGTGGTATTCATCGACGATATTCTTATTTATTCTCAGAGTGATAGCGATCATGAACAACAT
CTAAGGTTGGTGCTACAGAAGTTGCGAGATAATCAACTCTATGCTAAGTTCAGTAAGTGCGAGTTTTGGA
TTGACAAGGTGTCATTCCTTGGACATATTATTTCTAATGGAGGAATAGCAGTGGATCCTGCTAAAGTGAA
GGAGATAATGGAGTGGAGAGTGCCCACTACAGTTACTGAGATTCGGAGTTTCTAAGGGCTTGCAGGATA
TTATCGGAGGTTTATTGAAGGATTCTCTAAGATTGCCAAGCCTATGACATCGCTTCTGGAGAAAGGAAGA
GAATTTAAGTGGGATGAGAAGTGCCAAGAAAGCTTTGATCAATTGAAGGAGAGATTGATGTCACCTCCA
GTGTTGGTTATGCCAGACTTACAGAAAGGATTTGACATTTATTGTGATGCATGTGGCCAAGGACTTGGAT
GTGTGCTCATGCAAGAAGGTCATGTGATTGCCTATGCATCTCGTCGGTTGCGGAAACATGAATTGAACTA
CCCCACTCATGATTTGGAATTGGCAGCAGTTGTGCATGCACTTAAGATTTGGAGACACTACATTATGGGA
ACTAAGTGTCAAGTGTACATGGATCATAAGAGTTTGAAGTATATATTCACTCAGAAGGACCTCAACCTTA
GGCAACGCCGTTGGTTGGAGCTTATTAAGGATTATGATTTGGAGATTCACTATCATCCGGGCAAGGCGA
ATTTGGTCGCAGATGCCTTGAGTCGGAAAGAACATGTTCACTCCGCTTTTGTTGAAGGAGTCACCATCGA
GTTGGAACCTACCTTGGAGCAAGAGATTCGTAAAGGTCAGGTTGGTGATGCTAAAATTCAAGAGATTAA
GGATCTGATTACGGAAGGTCGAGTCCCAGAATTCACGGAGGATGAGCAAGGTACTATATGGTTCAAGGA
TCGAATATGTGTTCCAGAGATTGATAGCCTTCGTGAGACGATATTGAAAGAAGCTCATGATTCTGATTATT
CTATCCACCCTGGTAGCACTAAGATGTATCAAGATTTGAAGCAAAAGTATTGGTGGTATGGACTGAAAAG
AGATGTTGCTGCTCATGTGGCTAAGTGTGATGTGTGTCAAAGAGTCAAGGCCGAACATCAAAGACCAAC
TGGACTATTGCACCCATTGAAGATACCTGAGTGGAAATGGGAAGAAATTGGTATGGATTTTATTGTTGGA
CTACCCCGCACCCCTGCTGGATATGACTCTATTTGGGTGATTGTAGACAGATTGACAAAAGTAGCTCATTT
CATACCCGTGAGGACTAATTATACGGGAGCCAAGCTAGCAGAATTGTATATGACTCGGATAGTTTGTTTA
CATGGTGTGCCTAAAAAGATCATGTCGGATCGAGGATCACAGTTTACTTCTCGGTTTTGGAAGAAGTTGC
ATGAGAGTTTGGATACAAAGTTGAATTTTAGTTCAGCCTACCATCCCCACACTGATGGACAGACAGAGAG
AACTAACCAAGTATTGGAAGATATGTTGAGAGCTTGTGCCCTTCAGCATGGTAGTAGTTGGGACAAGAG
TCTACCTTGTGCTGAGTTCTCCTATAATAATAGCTACCAGGCCAGTCTGAAGATGTCACCATTTGAAGCTT
TGTATGGCAGAAAGTGCAGGACTCCACTGTACTGGGATCAAACTGGTGAAAGGCAGTTGTTTGGGCCTG
AGATTATACAAGAAGCAGAGGAACAAGTTCAGCAAATAAGGGAGAATTTGAGGGTAGCTCAAACCAGA
CAAAAGAGCTACGCAGATAATAGAAGAAGACCACTGGAATTTGAGAAAGGTGATTATGTGTATCTCAAG
ATGTCACCACTTCGTGGTATGAGAAGATTTAAAGTCAAAGGAAAGTTGTCCCCTCGCTTTATTGGACCCTT
CTTAATCTTGAAGCGAGTGGGAGAGGTCGCATATCAATTGGAGTTACCGAATCATCTTGCGGATGTTCAT
GATGTTTTTCATGTATCTCAGCTGAAGAAATGCCTCAGGGTACCTGAGGAACAATTATCAATGGAGGACC
TTAGTGTTCAGGATGATCTAACTTATGCAGAGTACCCCATCAAGATTCTTGATACACTGACTCGAGTCACG
AGGAATAAGGTTATAAAGATGTGCAAAGTTCAATGGAGTTACCACGGTGAAGATGAAGCAACTTGGGA
AAGAGAAGAAGAGCTTCGTATAGATTTTCTCCACCTTTTCCCTAGATCTTCTTAAATCTCGAGGACGAGAT
TTTTGTTAAGGGGGTAGGATTTGTAATACCCAACTTGTAAGAGATAATATAAAAAGAGAGAAAAATTATT
TTTCCCTTATTTATGTGTGTGTGATGCATCTATTTACCATCATATGTGAACACTATATTAAAAAAAACAATA
AATAATAAAAGATGTGTCGCTAAATTGTGCATCATGCTGGACTTTAATTTATGTGCATTTTGTGGCAATAT
AAAAATAATGTGAAAATAGTAAAGTCCAAAGGGGAATCTAAGATCTAATAAAATTCAACAATTTAGAAA
AAGAGAAGAAGGAAACAATATTATGTGAGTGAAATGAGTATAAAAAATATATATAATAAAATAAACTAT
TTCTGTAATGGATTTGAAGGTGTACATTGGACTCAAATGTGCAACTCCTAAAAGGATTCAAATTCAAATTT
GGAATTTGAACTAGAAAGAAGACAAAATAAGAGAAAGAGTACATATATATATATATATATATATCATTTC
AGGATAGTACTTTTATTTGAGCACTTGAACCCTGTGTGAAACCCATAAACAAATTTGAATTCAAATTATAG
ATTGGAGTGGGAGAATCAAAACAGGAATTAAAAAAATGAGAAAAGGAATTCCTGGCGCCTTGGGCCGA
ATTCTTAGTCCCGCGGCCCATTTTCCTCCACCCAGCACAGCCCAGCTCCTACTCTCCTTGCGTCACACTGAC

ATGTGGGGCCAGCATGGTCAGCCACCCGCGCGCTATCCTTCTCGCTGGCGGGTGGAGCCCACTTGTCGG
CACTCTCTTCTCCGTCGTAACGCCCGCGCGGAATTGCCGCTCGCGCTCGCACCTCACCCTGTCCCTGGGAG
GCTGGGATGCGGACTCGCGTGCCAGCCGCCGATCGTGAACGTGCTATCGCTGCGCTGTGGGGCCCGTTC
TCCTGTGCACGCCCCATTTCCGCAACAGCTCGCGCGGGAGTCGCCGTCTTCGTCGGGGATCGTGGGCGA
CTTCTCTTCCCCCCAAATCCCGTGAGTCCGCGCCGCAATCAACTCGCCTGGCTGTTATATGGGAGCTTAGG
TCGCGATCTCCTCGGCCTTCCACTCT

B73

>chr4:203128924-203133581

TCACTAGGCGCAGGCGCGCGCGTGATGTCCGGCGAGCAACCGAGCGCGCGGCCGCAGAGCAGGCAGCA
GCGCATGCTGCAAATCAGCCACCAATGGCAAGTGGCTGAGATTAGGAAAGAGGGCTAAAATCAGGTGC
TGCAAATCAGCCACAGATAGCAGGGGCTGAGATTAGGAAAGAGGGATAAGAATAGGAAAGAAGCAATC
AATCAGGCTAGAGAAAATCAAGGAGACTTGCTCCATTTGTAGGTGGCCGACGGCCATATATACCCAGCT
ATGCTGCACGGTTATAGCAATAAGGAATTTGTCTCTTCTTCCACCTCCAGCCTCACAAGCCTGCGGTTAGG
GGGGCATAACCCCAACCGGAGAGGACGGCCGACGGCTACGAGCTACGTAGCCGTGGTCCGGCAAGTCG
GGGATTCCAGGTCCTTGACAGCTTGGTATCACGAGTCCGGCGATCTTCGTCCAGCAATTCCCAGTCCACC
GAACCCCCCACCAGCCGCACAACACCACCATCTCCCACCCGCTCCTCTGCCCCCAGCACCATGTCGGAGTA
CACCCTTGCCGACGTCGTGAAGATGATGCAGGAGATGCAGGCCGACATCTCCGGCCTCAAAGAGAAGTC
CGCGTCGTCGTCCGCCGGCGGCACCATGCAGGAACGCCCTCGCGACACCGACCGCCTGCCTAAGTTCCA
GAAATGGGACTTCCCTCGGTTTGATGGCAAGACCGATCCTCTCCTCTTCATCAACAAGTGTGAGTCATACT
TTCGCCAACATCGTTCCTCGCCAGAGGAGCAAGTGTGGATGGCCTCATACAGTCTCGAAGGCACCGCGC
AACTATGGTACATGCAGCTGCAAGAGGATGACGGCACCCCGTCCTGGCGCCGGTTCAAGGAGCAGCTCA
ACCTGCGCTTCGGGCCGACCAGCCGGTCCGCGCCTTTCTTCGAGCTCACCGAATGCCGCCGCACCGGCTC
GGTCGAAGAGTACTCCAACCGTTTCCAGGAACTCCTACCCCACGCCGGTCCGATGGAGGAGGCGCAGCG
CGTCCAGCTCTACACGGGGGGGTTGCTGCCGCCGCTCAGTCACTCGGTGCGCATCCTACACCCCGACACT
CTCAGCGCCGCGATGAGCCTGGCGCGGGAGGTGGAACAGCTGGAGATCGCCCGCGTCCAGGCGGCGAC
AAGACCGGCAGCGCGGGGTATTCTGCCGGCGCCCCCTCCACGACCAGCCGCTCCGACACTACCGGCGCC
ACCTCCTCCCTTGGCCCTGCCAGCGCCGCCGGTGGGTGCAGCACACGCGCGCGGCGAGGGCGCTCGCCG
CCTGTCACCCGAGGAGATGGCGGATCGTCGTCGGCAAGGCTTATGTTTCAATTGTAACGAGAAGTACTCT
CGGGGCCACAATCGTTTTTGCAGGCGTCTCTTTTTTGTGGATGGGGTGGAGATCGATGACACCGGGGTC
GACGGGGACACGGGCGCGGCCGAGCCCAACGCCCCGCTCTTCTCCCTGCAGGCGGTGGCCGGGGTGGC
GCCGACTGACACCATGCAGATCGCGGTCACCCTCGGTCCGGCCTCCCTGGTGGCGCTGCTCGACTCGGG
CAGTACACACAACTTCATCTCCGCGACGGCAGCACACCGGTCGGGCCTCCCCATCCAGCGCCGTCCACGT
CTCACGGCCATGGTGGCTAACGGCGAGCGGATCTCCTGTGCCGGCGTCCTCCGCAACGCGCCGCTCCTC
GTCGAGGGCGAACCATACCCGGCCGATCTCTTCGTCATGCCGTTGGCCGGGTACGACGTCGTCCTCGGC
ACGCGTTGGCTGGGGGCCCTGGGACCGATCGTGTGGGACCTCGCCACCCGCCGCATGACGTTTCAGCGC
CCAGGGCGCTCGGTCTCCTGGACGGGTACCCCACGCGCAGCGCAGCCGGCGGTGCGCACCATGACAGG
GGGGGATGCGTTGATGGAGGCTCTCCTGGACTCGTTTGGGGGGCTCTTCGCCGCTCCTTCGGGCTTGCCT
CCCAAGCGCGCCCACGACCATCGCATCCTCCTCAAGCCCGACGCATCGCCGGTGGCTGTCCGCCCGTACC
GCTACCCGGCTGCCCACAAGGACGAGCTGGAGCGTCAGTGCGCCGCGATGATCGAGCAAGGGATCGTG
CGCCGCAGCGACTCCCCGTTTTCGTCACCGGTTCTCTTGGTCAAGAAGCCCGATGGCTCCTGGCGTTTCTG
CGTCGACTACAGAGCCCTGAACGCCCTCACCATCAAAGACGCGTTCCCGATCCCGGTGGTAGAGGAGCT

GCTCGACGAGCTCCACGGCGCCAAGTTCTTCACTAAGCTCGACCTTCGGTCCGGCTACCATCAGGTGCGG
ATGCGGCCGGAGGATGTGCACAAGACCGCCTTCCGCACGCACGATGGCCTGTACGAGTTCCTCGTGATG
GCGTTCGGACTCTGCAACGCTCCGGCTACGTTCCAGGCGCTGATGAACGACGTGCTCCGACCATTCCTCC
GCCGTTTCGTCCTGGTATTCTTTGACGACATTCTAATTTACAGCAAAACCTGGGCCGATCACCTCCGCCAT
CTCCGGGCGGTTCTCGACGAGCTGCAGCGCCACCGGCTCTTCGTCAAGCGCTCGAAGTGCTCTTTCGGCG
CGCCGGCCGTCTCCTACCTCGGCCACATTATCTCGGCAGCCGGGGTGGCTATGGATCCGGCCAAGGTGG
CCGCCATCCGCGACTGGCCGCCACCTCGCTCGGTCCGGGCAGTTCGCGGTTTCCTCGGCCTTGCCGGGTA
CTACCGCCGCTTCGTCCACAACTACGGCACCGTGGCTGCCCCGTTGACGGCCCTCCTCAAGAAGGATGGC
TTCGTCTGGGACGACGCGGCTATGACAGCGTTCGCAGCACTCAAGGCAGCCGTGTCCTCCGCCCCGGTG
CTCGCCATGCCGGACTTCAGCAAGCCGTTCGTGGTGGAGTGCGACGCGTCCGCCACAGGGTTCGGCGGG
GTGCTTGTCCAGGAGGGACACCCGGTCGCGTTCTTCAGCCGGCCAATCGCTCCCCGGCATCGCGCCCTCG
CCGCCTATGAGCGCGAACTTATCGGCTTGGTCCAGGCCGTTCGCCATTGGCGGCCATACTTGTGGGGTCG
GCACTTCACCGTCAAGACTGACCATTTCAGTCTCAAGTACCTCCTCGACCAGCGGCTGGCCACCATCCCTC
AACACCATTGGGTGGGAAAGCTCCTGGGCTTCGACTTCGAGGTGGAGTACCGCTCGGGTGCCACGAACA
CCGTCGCCGACGCCCTGTCCCGAAAAGACACCGAGGAGGGCGCGCTGATGGCTATATCCGCACCCCGCT
TCGACTTCATCACGCGGCTACGCCAAGCCCAAGCTACGGACCCGGCCCTCGTCGCCATCTGCGAGGAAAT
TCGCGCGGGCACACGTACGGCGCCGTGGTCCATCATGGGCGGGATGGTCGCCATGCATGGGCGCCTCTA
CATCCCGGCATCGTCCCCTCTCCTGCAGGAAATAGTGGCTGCCATACACGACGATGGCCACGAAGGCGTC
CTACGCACCCTACATCGGCTGAAGCGCGACTTCCATTTTCCAAACATGCGCCGTGTTGTGCAGGATTTCGT
CAAGGCATGCGGCACGTGCCAAAAATACAAGTCTGATCACTTGCGGCCGGCCGGTCTTCTACAGCCCTTG
CCGATCCCCTCGGCGGTGTGGGCCGACATCGGCATCGATTTCATCGAGGCGCTGCCCAGCGCCCAAGGC
AAGACGGTCATCCTCACCGTCGTCGACCGCTTCAGCAAGTACTGCCATTTCATCCCGCTGGCGCACCCGTA
CACGGCAGAATCGGTGGCGCAGGCGTTCTTCGCCGAAGTGGTGCGGCTGCACGGCGTTCCGCAGTCCAT
CGTCTCGGACCGCGACGCCGTGTTCACTTCGGCCTTCTGGACAGAGCTGATGGAGCTCACGGGCACCAA
GCTCTTCATGTCCTCCGCCTTCCACCCCCAGACAGACGGCCAGTCGGAGGCAGCCAACCGCGTCATCGCC
ATGTACCTTCGCTGCTTCACCGGCGACCGCCCTCGACAGTGGCTCCGGTGGCTTCCGTGGGCGGAGTACA
CCTACAACACTGCCTTCCAGTCATCACTCCGCGACACGCCATTCCGGGTGGTGTACGGCCGCGACCCGCC
AACCATCCGATCCTATGAGCCGGGCGATGCACGCATTCCCGCGGTGGCGCAAGAGATGGAGGAGCGCA
GGCGCAGGCGCGCGCGTGATGTCCGGCGAGCAACCGAGCGCGCGGCCGCAGAGCAGGCAGCAGCGCA
TGCTGCAAATCAGCCACCAATGGCAAGTGGCTGAGATTAGGAAAGAGGGCTAAAATCAGGTGCTGCAA
ATCAGCCACAGATAGCAGGGGCTGAGATTAGGAAAGAGGGATAAGAATAGGAAAGAAGCAATCAATCA
GGCTAGAGAAAATCAAGGAGACTTGCTCCATTTGTAGGTGGCCGACGGCCATATATACCCAGCTATGCT
GCACGGTTATAGCAATAAGGAATTTGTCTCTTCTTCCACCTCCAGCCTCACAAGCCTGCGGTTAGGGGGG
CATAACCCCAACCGGAGAGGACGGCCGACGGCTACGAGCTACGTAGCCGTGGTCCGGCAAGTCGGGGA
TTCCAGGTCCTTGACACAAAC

>chr5:123190757-123194673

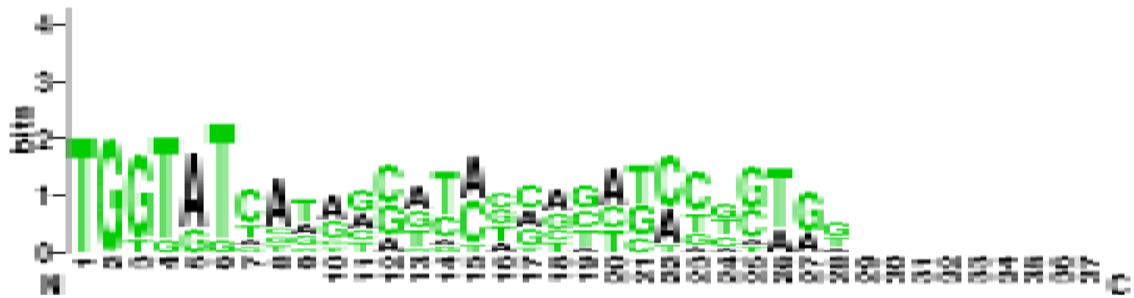
TGGTTCGATTGCTCCATGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGTTTTGCGTGGAGT
TGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAATAGTGGGCCGG
CCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAATAGAAACAGAAT
CCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTTGGCGCGGCTCA
CGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCATCTGGTATCGGATTCGGGGTCGGCTGCGT
CTTCATCCTCTTCCACGTGCTGTTCCTCGTCTTTGGCGGTTCGTGGGCGACTTCTTCTTCTGCTGGGTCCAT

CCAGTTCATCACTACGGCGCCCGCAACCAAACCATCGGTACAGACTCGTTTGGTGTTACAGGCTATGGCG
GATGGGGAGCGAGTTAGTCAGGAGAATTGGAGTCGTGTTATTGAGACGTTGGAGGGAGTGAGTAGCAG
GTTGAAGGATGTGGAGAGGGTGCAGCAACAAATCATGATGCGGACTGAATTAGCGGCATCAGTTGCTG
AACAGGCTGCGAAGGAGAGAAATGATTTGGGTCGTAAATTGGAGGAGACTGGAAGGGGAATGATATTT
GGGCTTTAGTGGTGTTGGAAGAGGTGTCTGCAGATCAGGGAGTACAGCAGCAGGGGGATCTTCAGCAG
TTACTTCAGGAGTTTGGAGATGTGTTTGAAGTGCCTTCTCAGTTACCTCCTGCTAGGGTGTATGATCATCA
TATTCCTTTGGTTCCTGGAGCAGTACCAGTGAATTCCAGACCTTATAGGTACTCTCCTCTTCACAAGGATG
AAATTGAGAGACAAGTGGCTGCTTTGTTGAAGGCTGGTCTTATTGTTCCTAGTGTAAGCCCGTTTGCTTCT
CCGGTATTGTTGGTTAAGAAGAAAGATAGCACTTGGAGGTTTTGTGTGGACTACAGGAAATTAAATTCTT
TGACCATTAAGAACAGATTTCCTATGCCGCTTGTAGAAGAGATTCTGGATGAACTGGCTGGTACTAAATA
TTTTTCTAGTCTGGATCTTACTGCTGGGTATCACCAGATTAGAATGGGGGAAGGTGATGAGCATAAAACA
GCCTTTAAAACCCATCAGGGTCATTACCAGTTCAGGGTCATGCCTTTTGGGTTAACCAATGCTCCTGCAAC
CTTTCAATGTGCTATGAATTCTGTTTTGGCTCCTTTTCTTCGGAAATTTGTCATGGTTTTCATTGATGACATT
TTGGTCTATAGTGCTACATGGGAGGAACACTTGGCTCATATTAAGATGGTGCTGCAGACTTTGAGAGAG
CACCAGTTCTATCTGAAGAGGAGTAAATGTGTGTTTGGGAGGAATGAGCTCATTTATTTGGGACATATTA
TCTCTCACACAGGAGTGGCCACTGATCCTGCTAAGACAACAGCCATGGAGAAATGGCCTGTTCCTCAGTC
AGTAACTGAGTTGAGGGGGTTTTTGGGTCTCACTGGGTACTATAGGAGATTTGTGAGACATTATGGGAT
TATTGCAAAACCTCTTACTAATCTGTTGAAGCATAAGGCTTTTCAGTGGAGTCCTGAGGCTGATCAGGCTT
TTCAAGCGCTGAAGAAGGCTATGCTTACTACTCCTGTGTTGGCCCTACCCAGGTTTGATGAGTTATTTGTT
GTGGAGACTGATGCTTGTGATGATGGGATAGGTGCTGTTTTGATGCAATGTGATAGACCAGTGGCATAT
CTAAGTAAAGCATTGGGTGCTAAGAATAAGATGCTATCCATTTATGAAAAAGAGTTTTTGGCTCTAATCA
TTGCAGTGGACAGGTGGAGGCAATATTTGCAGAGGGCTGAATTTGTGATTAAAACTGATCATCAGGCTT
TGAGTTTTCTGGATGATCAGCTTTTGCAATCTGATATGCAGAGGAAGGCTATGGCTAAGTTGATGGGGCT
TCATTTCAAAATTGTTTACAAAAAAGGCAAGGACAATGTGGCTGCCGATGCTCTATCCAGGATTGGTCAT
GTGATGCAGATTCACACGGTGACTGAGGTGAAACCAGCTTGGGTGCAAGAACTCCTGAATGCTTATGTG
ACAGATAAAGAGGCTCAAGATCTATTGGCTCAATTGGCCCTGGCCAGTCCTAATGAGCAAGGTTTTTCTT
TACATCAAGGTATTATTAGAAAAGGGGATCAAATCTGGATAGCTAACAATTCTGCCCTTCAAACTAAGTT
GATTTCAGCCCTGCATGATAGTGTGGTGGGAGGACATTCTGGTGGTGAGGCCACTTACCAAAGGATCAA
GAGGTTGTTCTGGTGGAAAGGGCTCAAGTCAGCCGTGTTTCAGTTTGTGCAGCAATGTGAGGTGTGTCA
GCATGCTAAGCATGAAAGGGTGCATTATCCGGGTCTCCTACAACCTTTACCAGTGCCTCAAGGGGCTTGG
CAGGATCTTACATTGGATTTCATTGAAGGTCTACCTAGATCTGAGGGATATGATACTATTTTGGTCGTGGT
GGATAGATATTCCAAATATGCTCATTTTTTTCCCCTCAAGCATCCTTTCACTGCTGTTGGAGTGGCACAGG
TGTTTCTTGACAATATTGTCAAGTTGCATGGGACACCAAAGTCTTTGGTTTCCGATAGAGACGAAGTGTTC
ACTAGTTTGTTTTGGAAATCTTTGTTTCAGGCTTTGGGGATCAAATTGGCTCTCACTACTGCATATCACCCT
CAAACTGATGGGCAATCAGAGCGAGTGAATCAGTGTTTGGAAATGTATCTTCGCTGTTCTATTCATGAGG
CACCTAATACTTGGAAGAAATGGTTGGCCCTTGCTGAATTTTGGTACAATACTTCATATCACACAGCTTTG
GGGTGTTCTCCATTCAAGGCCTTGTATGGTTATGACCCTGTGTTTGCTGCTGCTCCTATGCTTTCTGGGGT
TGCTCCTAGTTCTGTGACTGAGGTGTTGGCTGATAGACAGTTACATGCTAGTATGCTGAAGGAGAAGTTG
GCAGCTGCTCAAAATAGGATGAAACTCCAGGCTGATAAGCATAGATCTGACAGAGAGTTTCAGATTGGA
GATAGCGTTTTGGTAAAGCTTCAGCCTTATGTGCAGAAGTCGGTGGTTAGTCGACCCTGTCCTAAATTGG
CCTATAAGTTCTATGGTCCTTTTAAGGTGTTACAGAAGATTGGGATGGTTGCTTATAAGTTGGATTTACCA
GTTACATCTTTGATTCATCCAGTATTCCATGTATCTCAGTTGAAGGCTTTTACTCCCAATTATTCTCCGGTG
TTTTCTGATTTACCACAACTTGTTGATCTGTCTGTTAAAGAGATTATGCCATTGGCTGTGGTGGAAAGGAG
GATGGTGAAGAAGGGTAATCATCCTGTGGTGCAAGTAAGGGTGACTTGGTCCAATATGCCAGCTGAGGT

TACCACTTGGGAGGACTATGAAGTGTTGAAGAACCGGTTTCCTACAGCTCTTGATTGGGGACAATCAAGC
TCTTCCGGGGGGGGGGGGGGGGGCAAGTGTTAGTACTGCTACTGGTATGGCGATTTAATCTTTGTGCTGT
TTTGCGTGGAGTTGTAATTGGGCTGTTAAGCCGGGATTTGGGGTGTCAATGTGGCTAATTAGCTTTGTAA
TAGTGGGCCGGCCTTATTAGCTTGACTAAAAGCTGTAACCTGGAGTGGATCTGGAAATAAGAAAGAAAT
AGAAACAGAATCCCAACAAACTTATCTGTGAGTTTGAGCGCTTTTCTTCTGTTCCTCTCCTTTGCTGGATTT
GGCGCGGCTCACGTCGTCGAAGTTGCGCATCGGCGGCCACGGGCGGTGTCACTCCAGTGATGATGCATG
GT

APPENDIX B

PBS Logos

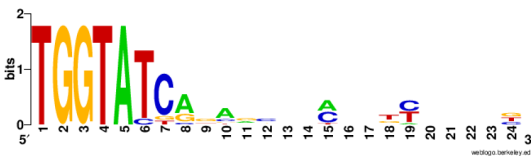


PBS Logo 5

PBS Logo 4

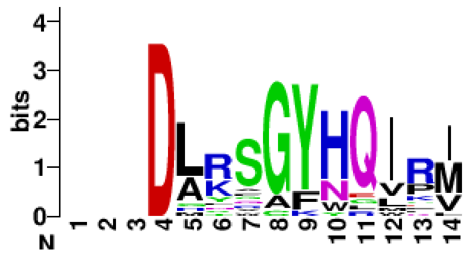
PBS Logo 3

PBS Logo 2

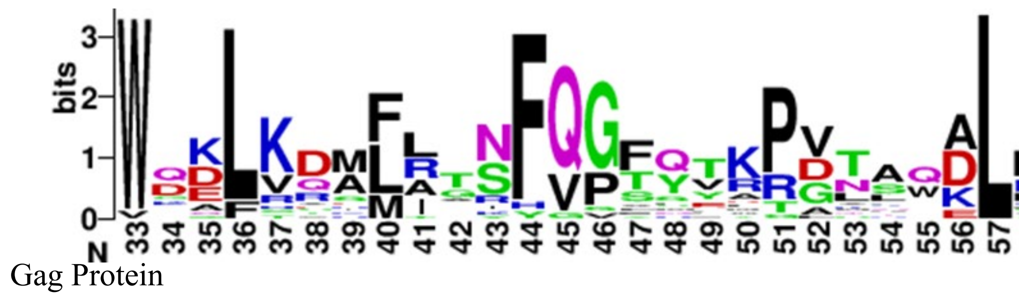


PBS Logo 1

RNase



Reverse Transcriptase



Gag Protein

APPENDIX C

```
import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;
import java.util.Arrays;
import java.util.HashMap;
import java.util.Scanner;
import java.util.regex.Matcher;
import java.util.regex.Pattern;

public class findRetrovirus_RNAse_GAG {
	static HashMap<String, String> codonTable = new HashMap<String, String>();
	static HashMap<String, String> origCandidates = new HashMap<String, String>();
	static HashMap<String, String[]> candidatesAmino = new HashMap<String, String[]>();
	static HashMap<String, String> candidates = new HashMap<String, String>();
	final static int NUM_RETRO_PATTERNS = 3;
	final static int NUM_RNA_PATTERNS = 2;
	final static int NUM_GAG_PATTERNS = 1;
//	final static int NUM_INT_PATTERNS = 1;

	public static void main(String[] args) throws IOException {
		initCodonToAminoMap();
		loadChrMapTEMP(args);
		Pattern[] retroPatterns = new Pattern[NUM_RETRO_PATTERNS];
		Matcher[] retroMatchers = new Matcher[NUM_RETRO_PATTERNS];
		Pattern[] RNAPatterns = new Pattern[NUM_RNA_PATTERNS];
		Matcher[] RNAMatchers = new Matcher[NUM_RNA_PATTERNS];
		Pattern[] GAGPatterns = new Pattern[NUM_GAG_PATTERNS];
		Matcher[] GAGMatchers = new Matcher[NUM_GAG_PATTERNS];
		retroPatterns[0] = Pattern.compile("NAPA");
		retroPatterns[1] = Pattern.compile("KTAF");
		retroPatterns[2] = Pattern.compile("M.FGL");
		RNAPatterns[0] = Pattern.compile("DAS.........Q");
```

```
		RNAPatterns[1] = Pattern.compile("D..SR");
		GAGPatterns[0] = Pattern.compile("............W.........................................L");
		for (String key : candidatesAmino.keySet()) {
			for (String aminoSeq : candidatesAmino.get(key)) {
				retroMatchers[0] = retroPatterns[0].matcher(aminoSeq);
				retroMatchers[1] = retroPatterns[1].matcher(aminoSeq);
				retroMatchers[2] = retroPatterns[2].matcher(aminoSeq);
				RNAMatchers[0] = RNAPatterns[0].matcher(aminoSeq);
				RNAMatchers[1] = RNAPatterns[1].matcher(aminoSeq);
				GAGMatchers[0] = GAGPatterns[0].matcher(aminoSeq);
				for (Matcher retroMatcher : retroMatchers)
					if (retroMatcher.find()) {
						for(Matcher RNAMatcher : RNAMatchers) {
							if(RNAMatcher.find()) {
								for(Matcher GAGMatcher: GAGMatchers)
{
									if(GAGMatcher.find()) {

	candidates.put(key, origCandidates.get(key));
									}
								}
							}
						}
					}
			}
		}
//		System.out.println(candidates.toString());
//		System.out.println(candidates.keySet().toString());
		writeNewCandidates(args);
	}

	private static void writeNewCandidates(String[] args) throws IOException {
		System.out.println(args[0]);
		String outputDirectory = args[0].replace("findPBS-sameleftright", "retrovirus-RNAase-GAG-CandidateLTRs-
");
		System.out.println(outputDirectory);
		File file = new File(outputDirectory);
		file.createNewFile();
		FileWriter writer = new FileWriter(file);
```

```
		for (String key : candidates.keySet())
			writer.write(key + "\n" + origCandidates.get(key) + "\n\n");

		writer.close();
	}

	public static void loadChrMapTEMP(String[] args) throws IOException {
		File chrs = new File(args[0]);
		Scanner fileReader = new Scanner(chrs);
		while (fileReader.hasNextLine()) {
			String name = fileReader.nextLine();
			String chr = fileReader.nextLine();
			fileReader.nextLine();
//			System.out.println(name + " -> " + chr);
			origCandidates.put(name, chr);
		}
		for (String key : origCandidates.keySet()) {
			candidatesAmino.put(key, convertBoth(origCandidates.get(key)));
//			System.out.println(key + " ---> \n" + Arrays.toString(convertBoth(candidates.get(key))));
		}
		fileReader.close();
	}

	public static void initCodonToAminoMap() {
		codonTable.put("TTT", "F");
		codonTable.put("ATT", "I");
		codonTable.put("GTT", "V");
		codonTable.put("TTC", "F");
		codonTable.put("CTC", "L");
		codonTable.put("ATC", "I");
		codonTable.put("GTC", "V");
		codonTable.put("TTA", "L");
		codonTable.put("CTA", "L");
		codonTable.put("ATA", "I");
		codonTable.put("GTA", "V");
		codonTable.put("TTG", "L");
		codonTable.put("CTG", "L");
		codonTable.put("ATG", "M");
```

```
codonTable.put("GTG", "V");
codonTable.put("TCT", "S");
codonTable.put("CCT", "P");
codonTable.put("ACT", "T");
codonTable.put("GCT", "A");
codonTable.put("TCC", "S");
codonTable.put("CCC", "P");
codonTable.put("ACC", "T");
codonTable.put("GCC", "A");
codonTable.put("TCA", "S");
codonTable.put("CCA", "P");
codonTable.put("ACA", "T");
codonTable.put("GCA", "A");
codonTable.put("TCG", "S");
codonTable.put("CCG", "P");
codonTable.put("ACG", "T");
codonTable.put("GCG", "A");
codonTable.put("TAT", "Y");
codonTable.put("CAT", "H");
codonTable.put("AAT", "N");
codonTable.put("GAT", "D");
codonTable.put("TAC", "Y");
codonTable.put("CAC", "H");
codonTable.put("AAC", "N");
codonTable.put("GAC", "D");
codonTable.put("CAA", "Q");
codonTable.put("AAA", "K");
codonTable.put("GAA", "E");
codonTable.put("CAG", "Q");
codonTable.put("AAG", "K");
codonTable.put("GAG", "E");
codonTable.put("TGT", "C");
codonTable.put("CGT", "R");
codonTable.put("AGT", "S");
codonTable.put("GGT", "G");
codonTable.put("TGC", "C");
codonTable.put("CGC", "R");
codonTable.put("AGC", "S");
```

```
		codonTable.put("GGC", "G");
		codonTable.put("CGA", "R");
		codonTable.put("AGA", "R");
		codonTable.put("GGA", "G");
		codonTable.put("TGG", "W");
		codonTable.put("CGG", "R");
		codonTable.put("AGG", "R");
		codonTable.put("GGG", "G");
		codonTable.put("CTT", "L");
		codonTable.put("TAA", "-");
		codonTable.put("TAG", "-");
		codonTable.put("TGA", "-");
	}

	public static String codonToAmino(String codon, boolean print) {
		if (print) {
//			System.out.println(codon + "->" + codonTable.get(codon));
//			String reversecodon = new StringBuilder(codon).reverse().toString();
//			System.out.println(reversecodon + "-->" + codonTable.get(reversecodon));
		}
		return codonTable.get(codon);
	}

	public static String codonToAmino(String codon) {
		return codonTable.get(codon);
	}

	public static String[] convertBoth(String codonSeq) {
		String[] out = new String[6];
		String genestrand1 = codonSeq;
		String genestrand2 = codonSeq.replace('A', 'H').replace('T', 'A').replace('H', 'T').replace('C', 'H')
				.replace('G', 'C').replace('H', 'G');
		String[] first = codonSeqToAminoSeq(genestrand1);
		String[] second = codonSeqToAminoSeqAntiPair(genestrand2);
		return new String[] { first[0], first[1], first[2], second[0], second[1], second[2] };
	}

	public static String[] codonSeqToAminoSeq(String codonSeq) {
```

```
        String aminoSeqF1 = "";
        String aminoSeqF2 = "";
        String aminoSeqF3 = "";
        for (int i = 0; i + 2 < codonSeq.length(); i += 3) {
            aminoSeqF1 += codonToAmino(codonSeq.substring(i, i + 3));
            if (i + 3 < codonSeq.length())
                aminoSeqF2 += codonToAmino(codonSeq.substring(i + 1, i + 4));
            if (i + 4 < codonSeq.length())
                aminoSeqF3 += codonToAmino(codonSeq.substring(i + 2, i + 5));
        }
        return new String[] { aminoSeqF1, aminoSeqF2, aminoSeqF3 };
    }

    public static String[] codonSeqToAminoSeqAntiPair(String codonSeq) {
        String aminoSeqF4 = "";
        String aminoSeqF5 = "";
        String aminoSeqF6 = "";
        codonSeq = new StringBuilder(codonSeq).reverse().toString();
        for (int i = 0; i + 2 < codonSeq.length(); i += 3) {
            aminoSeqF4 += codonToAmino(codonSeq.substring(i, i + 3));
            if (i + 3 < codonSeq.length())
                aminoSeqF5 += codonToAmino(codonSeq.substring(i + 1, i + 4));
            if (i + 4 < codonSeq.length())
                aminoSeqF6 += codonToAmino(codonSeq.substring(i + 2, i + 5));
        }

        return new String[] {aminoSeqF4, aminoSeqF5, aminoSeqF6};
    }
}
```