

Generating Hemi-NeSTR input file from the UCSC Genome Browser

- 1) Access the Feb 2009 hg19 (note: important!) assembly in UCSC Genome Browser at: <https://genome.ucsc.edu/cgi-bin/hgGateway>
- 2) Type in your gene name of interest in the “search term” window (ex. COL3A1) and select the relevant gene from the list in the popped out pull-down menu

Screenshot of the UCSC Genome Browser interface showing the search results for the gene *COL3A1*.

The search results show the following details:

- Group: Mammal
- Genome: Human
- Assembly: Feb. 2009 (GRCh37/hg19)
- Position: chr3:187,417,489-187,417,641
- Search term: col3a
- Description: COL3A1 (Homo sapiens collagen, type III, alpha 1 (COL3A1), mRNA)

Below the search results, there is a link to reset browser settings and buttons for track search, add custom tracks, track hubs, and configure tracks and display.

Human Genome Browser – hg19 assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061,RH80175 15q11;15q13 rs1042522;rs1800370	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175, or chromosome bands 15q11 to 15q13, or SNPs rs1042522 and rs1800370. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.
D16S3046 AA205474 AC008101 AF083811 PRNP NM_017414 NP_059110	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well. Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17 Displays region of clone with GenBank accession AC008101 Displays region of mRNA with GenBank accession number AF083811 Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP Displays the region of genome with RefSeq identifier NM_017414 Displays the region of genome with protein accession number NP_059110
pseudogene mRNA homeobox caudal zinc finger kruppel zinc finger huntington	Lists transcribed pseudogenes, but not cDNAs Lists mRNAs for caudal homeobox genes Lists many zinc finger mRNAs Lists only kruppel-like zinc fingers Lists candidate genes associated with Huntington's disease

Homo sapiens
(Graphic courtesy of [CGPS](#))

3) Click "submit"

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

group	genome	assembly	position	search term
Mammal	Human	Feb. 2009 (GRCh37/hg19)	chr2:189839099-189877472	COL3A1 (Homo sapiens collagen, type III, alpha 1 (COL3A1))
<input type="button" value="submit"/> 				

[Click here to reset the browser user interface settings to their defaults.](#)

[track search](#) [add custom tracks](#) [track hubs](#) [configure tracks and display](#)

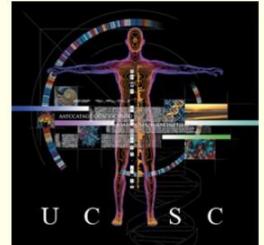
Human Genome Browser – hg19 assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061;RH80175 15q11;15q13 rs1042522;rs1800370	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175, or chromosome bands 15q11 to 15q13, or SNPs rs1042522 and rs1800370. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.
D16S3046 AA205474 AC008101 AF083811 PRNP NM_017414 NP_059110	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well. Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17 Displays region of clone with GenBank accession AC008101 Displays region of mRNA with GenBank accession number AF083811 Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP Displays the region of genome with RefSeq identifier NM_017414 Displays the region of genome with protein accession number NP_059110
pseudogene mRNA homeobox caudal zinc finger kruppel zinc finger huntington	Lists transcribed pseudogenes, but not cDNAs Lists mRNAs for caudal homeobox genes Lists many zinc finger mRNAs Lists only kruppel-like zinc fingers Lists candidate genes associated with Huntington's disease



Homo sapiens
(Graphic courtesy of CGPS)

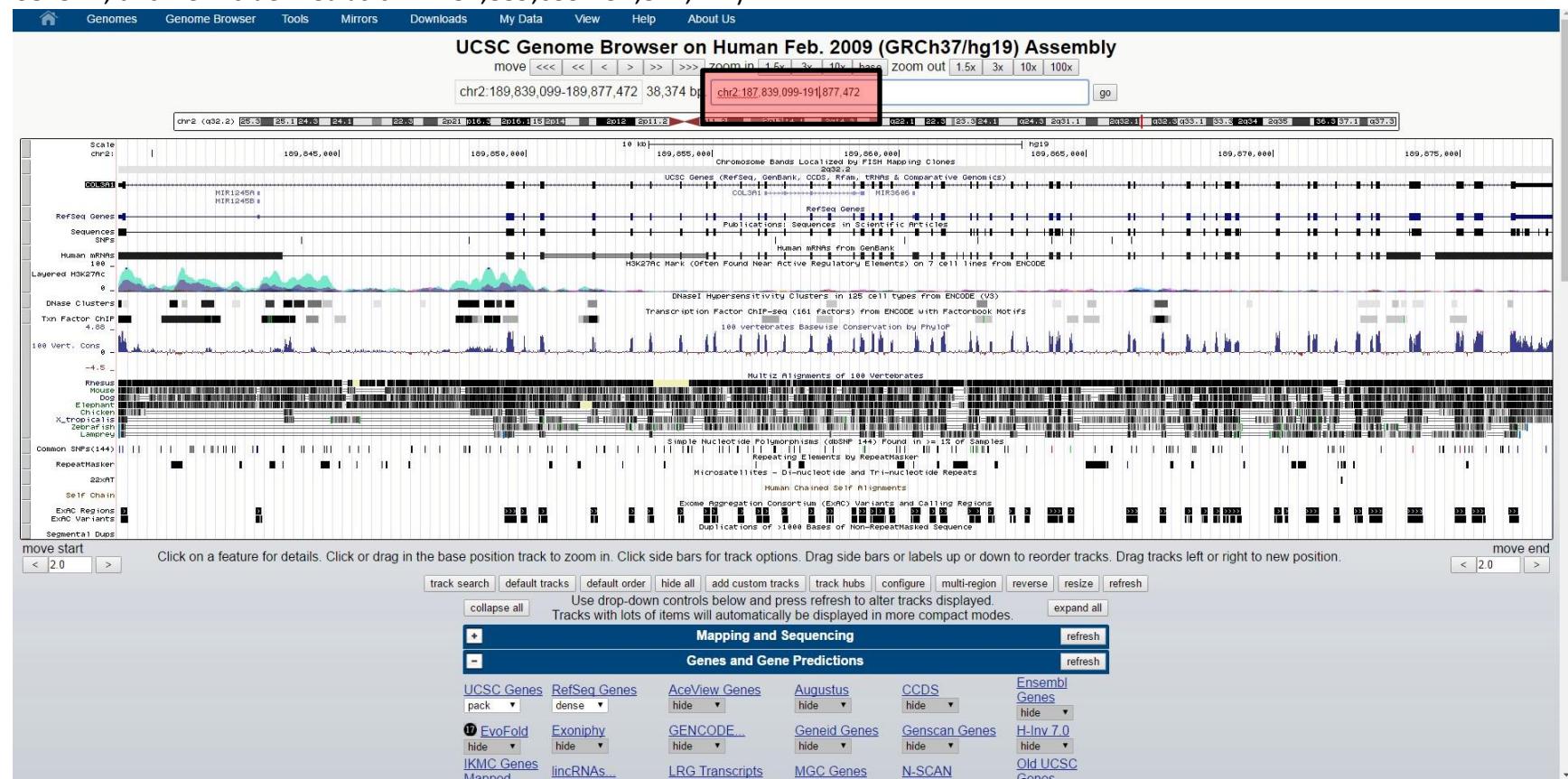
- 4) In the newly opened webpage, scroll down to the “Repeats” section at the bottom of the page and change the “microsatellite” pull-down to “full”; then click “refresh”

The screenshot shows the UCSC Genome Browser interface with the following sections visible:

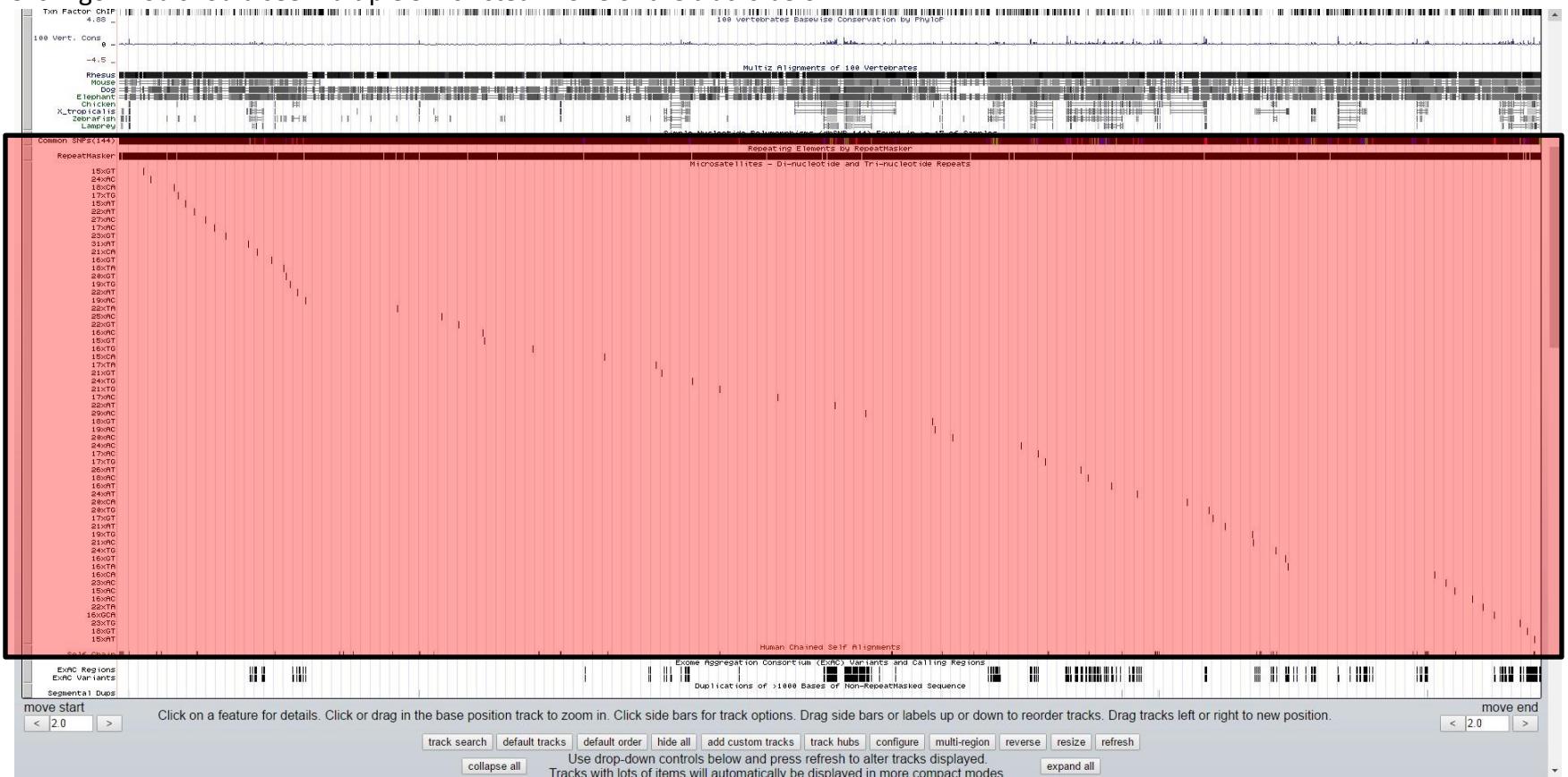
- Regulation:** Includes options like ENCODE Regulation, CD34 DnaseI, CpG Islands, ENC Chromatin, ENC DNA Methylation, ENC DNase/FAIRE, etc.
- Comparative Genomics:** Includes Conservation, Cons Indels, Evo Cpg, GERP, phastBias, and various chain/network options.
- Neandertal Assembly and Analysis:**
- Denisova Assembly and Analysis:**
- Variation:**
- Repeats:** This section is highlighted with a red box and a red circle around the "Microsatellite" dropdown. The dropdown is set to "full". To its right is a "refresh" button.

- 5) At the top of the refreshed webpage, click on the chromosomal coordinate window so that this text appears in the search window to the right. Adjust the view to show 2Mb (2,000,000 bases) downstream and upstream of the gene sequence (for

COL3A1, this view is defined as chr2:187,839,099-191,877,472)

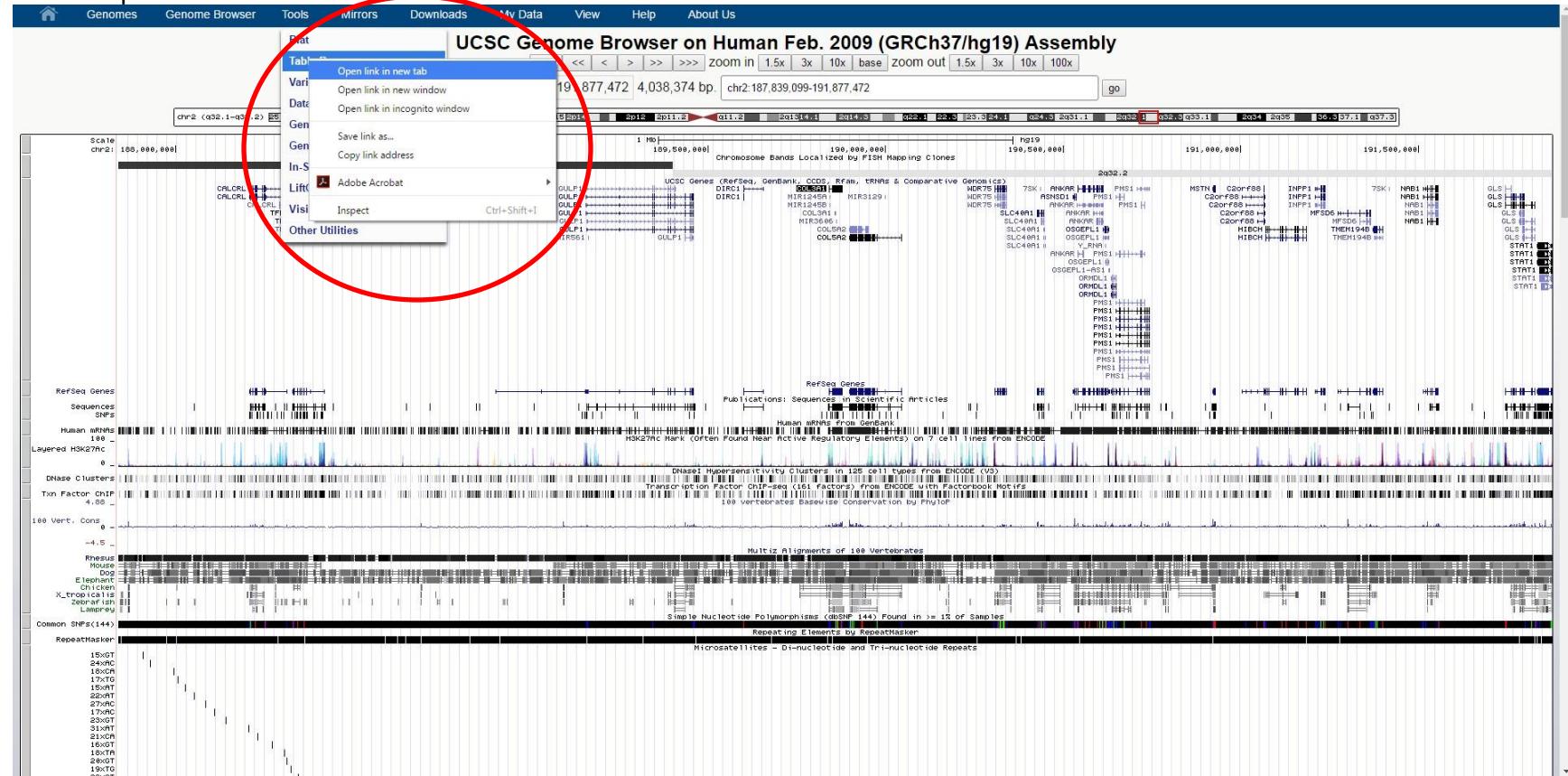


6) Click "go". You should see multiple STRs listed in one of the tracks below



- 7) Scroll to the top of the webpage and click on “Tools” then right-click “Table Browser” from the opened pull-down menu.

Select “Open link in new tab”



- 8) Select “Repeats” in the pulldown menu next to “group:”. Select “Microsatellite” in the pulldown menu next to “track:”. Be sure to select the radiobox for “position” in the “region:” section (otherwise, your computer may download all STRs in the human genome, which is not what you want). Select “sequence” in the pulldown menu next to “output format:”. Type a

name for your STR sequence file with a .txt extension (for example, col3a1.txt). Click “get output”

The screenshot shows the UCSC Table Browser interface. At the top, there's a navigation bar with links to Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. Below the navigation bar is a section titled "Table Browser". The main area contains several input fields and buttons:

- clade:** Human
- genome:** Human
- assembly:** Feb. 2009 (GRCh37/hg19)
- group:** Repeats (highlighted with a red box)
- track:** Microsatellite (highlighted with a red box)
- region:** position ch2:187839099-191877472 (radio button selected)
- identifiers (names/acccessions):** col3a1.txt (highlighted with a red box)
- filter:** create
- intersection:** create
- correlation:** create
- output format:** sequence (highlighted with a red box)
- output file:** col3a1.txt (highlighted with a red box)
- file type returned:** plain text (radio button selected)
- get output** button (highlighted with a red box and circled)
- summary/statistics** button

Below the form, there's a note: "To reset all user settings (including custom tracks), [click here](#)".

Using the Table Browser

This section provides brief line-by-line descriptions of the Table Browser controls. For more information on using this program, see the [Table Browser User's Guide](#).

- clade:** Specifies which clade the organism is in.
- genome:** Specifies which organism data to use.
- assembly:** Specifies which version of the organism's genome sequence to use.
- group:** Selects the type of tracks to be displayed in the track list. The options correspond to the track groupings shown in the Genome Browser. Select 'All Tracks' for an alphabetical list of all available tracks in all groups. Select 'All Tables' to see all tables including those not associated with a track.
- database:** (with "All Tables" group option) Determines which database should be used for options in table menu.
- track:** Selects the annotation track data to work with. This list displays all tracks belonging to the group specified in the group list.
- table:** Selects the SQL table data to use. This list shows all tables associated with the track specified in the track list.
- describe table schema:** Displays schema information for the tables associated with the selected track.
- region:** Restricts the query to a particular chromosome or region. Select genome to apply the query to the entire genome or ENCODE to examine only the ENCODE Pilot regions. To limit the query to a specific position, type a chromosome name, e.g. chrX, or a chromosome coordinate range, such as chrX:100000-200000, or a gene name or other id in the text box. You can select multiple genomic regions by clicking the "define regions" button and entering up to 1,000 regions in a 3- or 4-field BED file format.
- lookup:** Press this button after typing in a gene name or other id in the position text box to look up the chromosome position
- identifiers** (selected tracks only): Restricts the output to table data that match a list of identifiers. For instance RefSeq accessions for the RefSeq track. If no identifiers are entered, all table data within the specified region will be displayed.

- 9) **Optional:** You may wish to add custom short unique names to your STR sequences for downstream applications. In this case, you may skip to step 13) for a batch STR naming method.
- 10) Type "400 "in both windows and make sure that "Mask repeats" is unchecked. Click “get sequence”. This will download your desired STRs + flanking sequences into a txt file on your computer. Note the path of this file because it is your input fore the

Hemi-NeSTR program

The screenshot shows the Hemi-NeSTR website interface. At the top, there is a navigation bar with links: Home, Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. Below the navigation bar, the title "Microsatellite Genomic Sequence" is displayed. Underneath the title, there is a section titled "Sequence Retrieval Region Options:" which includes two input fields: "Add 400 extra bases upstream (5') and 400 extra downstream (3')". A note below these fields states: "Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome." Below this note is another section titled "Sequence Formatting Options:" containing several radio buttons and checkboxes. One checkbox is checked, and it says "Mask repeats: to lower case to N". There is also a link "Reverse complement (get '-' strand sequence)". At the bottom of this section are two buttons: "get sequence" and "cancel". A red circle highlights the "get sequence" button, and a black arrow points to it from the left.

- 11) Access the Hemi-NeSTR website. In step 3 “Paste your sequence(s)...” click “Choose File”. Navigate to the file you just downloaded (for example col3a1.txt) and click “Open.”
- 12) You may now click “Submit Job” if all other design parameters are suitable for your project design.

13) Perform step 8) as described above except leave the box next to “output file:” blank

The screenshot shows the Table Browser interface with the following settings:

- clade: Mammal
- genome: Human
- assembly: Feb. 2009 (GRCh37/hg19)
- group: Repeats
- track: Microsatellite
- table: microsat
- region: position chr2:187839099-191877472
- identifiers (names/acccessions): (empty)
- filter: (empty)
- intersection: (empty)
- correlation: (empty)
- output format: sequence
- output file: (highlighted with a red box)
- file type returned: plain text
- Send output to: Galaxy, GREAT, GenomeSpace
- get output
- summary/statistics

To reset all user cart settings (including custom tracks), [click here](#).

Using the Table Browser

This section provides brief line-by-line descriptions of the Table Browser controls. For more information on using this program, see the [Table Browser User's Guide](#).

- clade:** Specifies which clade the organism is in.
- genome:** Specifies which organism data to use.
- assembly:** Specifies which version of the organism's genome sequence to use.
- group:** Selects the type of tracks to be displayed in the track list. The options correspond to the track groupings shown in the Genome Browser. Select 'All Tracks' for an alphabetical list of all available tracks in all groups. Select 'All Tables' to see all tables including those not associated with a track.
- database:** (with "All Tables" group option) Determines which database should be used for options in table menu.
- track:** Selects the annotation track data to work with. This list displays all tracks belonging to the group specified in the group list.
- table:** Selects the SQL table data to use. This list shows all tables associated with the track specified in the track list.
- describe table schema:** Displays schema information for the tables associated with the selected track.
- region:** Restricts the query to a particular chromosome or region. Select genome to apply the query to the entire genome or ENCODE to examine only the ENCODE Pilot regions. To limit the query to a specific position, type a chromosome name, e.g. chrX, or a chromosome coordinate range, such as chrX:100000-200000, or a gene name or other id in the text box. You can select multiple genomic regions by clicking the "define regions" button and entering up to 1,000 regions in a 3- or 4-field BED file format.
- lookup:** Press this button after typing in a gene name or other id in the position text box to look up the chromosome position
- identifiers** (selected tracks only): Restricts the output to table data that match a list of identifiers. For instance RefSeq accessions for the RefSeq track. If no identifiers are entered, all table data within the specified region will be displayed.

- 14) Type “400” in both windows and make sure that “Mask repeats” is unchecked. Click “get sequence”. This will open your desired STRs + flanking sequences inside your web browser

Sequence Retrieval Region Options:

Add extra bases upstream (5') and extra downstream (3')

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

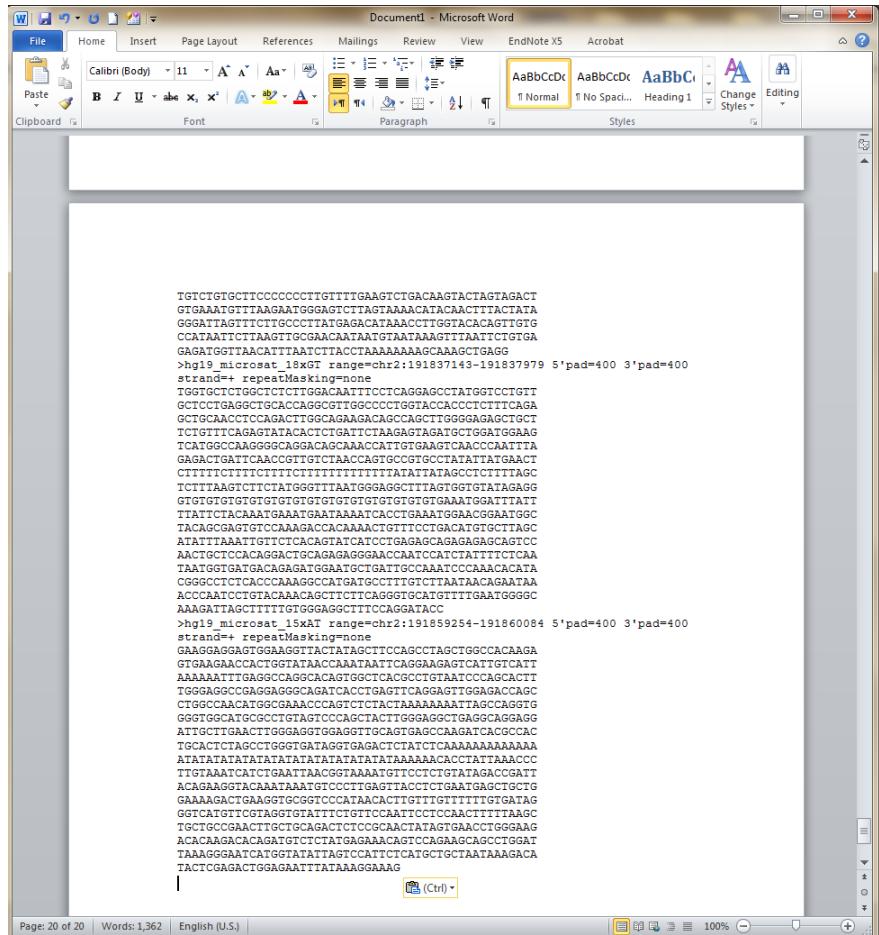
Sequence Formatting Options:

All upper case.
 All lower case.
 Mask repeats: to lower case to N
 Reverse complement (get '-' strand sequence)

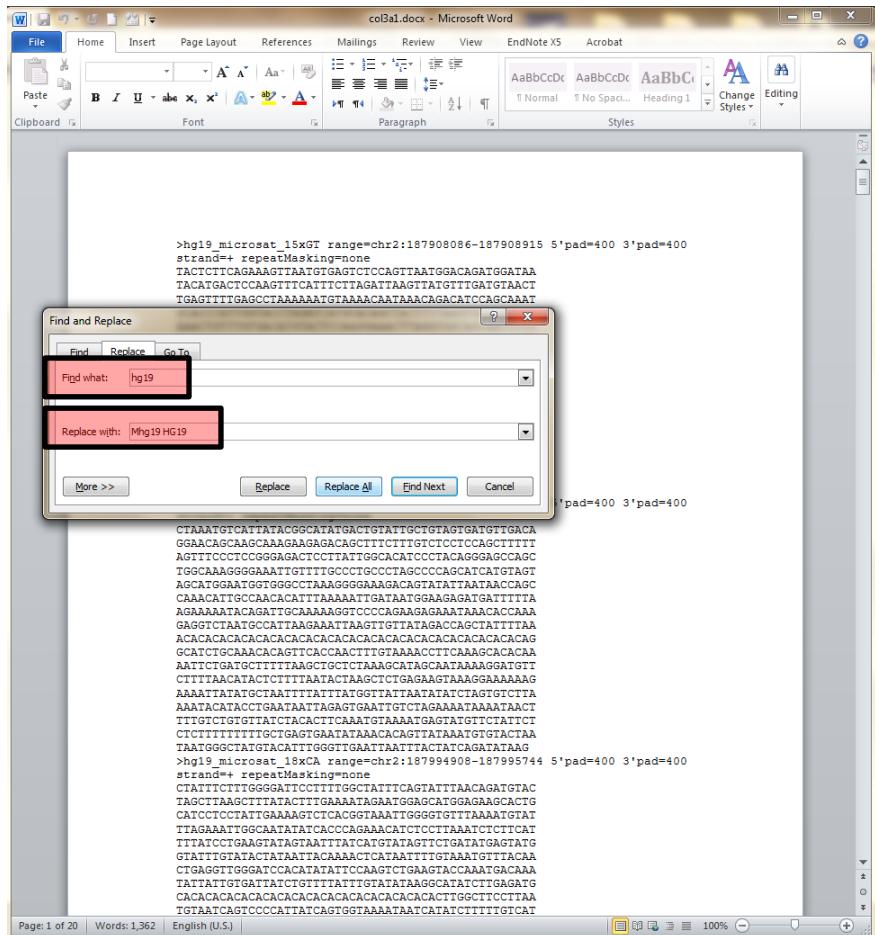
15) Press Ctrl+A to select all the text; then Ctrl+C to copy the text

```
>hg19_microsat_15xG1 range=chr2:187988895-187988915 5'pad=400 3'pad=400 strand+= repeatMasking=None
TACCTCTTCAAGAAAGTTAATGTGAGETCCAGTAAATGGACAGATGGATAA
TACATGACTCAGGTTTCAATTCTTAGATTTAAGTTATGTTGATGTAAC
TGAGTTTTGAGCTTAAAAAATGAAACAAACAGACATCGAGCAAAT
GCACCCATTGGTACCTAGATCATGACAGACTTAAAGTGTACACACA
AAACTGTTTGTGACATGTCACCAAGAAACATTAGTGGCAGTATTTC
AAATCTTGGAAATTGGTGAGATTCATCAAAAGATGGTACCTTAAAGCTT
TAGAGTGTGTCGTGTCCTGTTCAAGCAAGCTGGTAGCAAG
TTTTTGTATTAGGCTTCCTCTTCAAAATTTCCAACACATGCAAG
GTGTTGTTGTGTTGTTGTTGTTCAAAAGCAACACAGATTAT
TCAGAGTAACTTGTGAGTCAATGGGGATCTGAGCTTCAAGGAGGATCTGCT
CACTGACACTCCAACTCTCTGGTTAAAGGATTCTCTGGCTAGCCCTC
CTGAGTGGCTGGGATCACAGGACTCTGGCTGAAGATAAAAGTGAAC
AAATAATCATTTGTGAGTAAATACAAAGGGGAATGAGAATTTTGCAA
AATAAAAGAGGAACCTGGTAGGTCTAGTCAAGAAGTTCTGCA
AAAAGTAAAGTGAACCTGAGATTAAATAGAAATTTCGAAGGTGAGT
GGGGGAAAGGGTGTGAGGAGGAA
>hg19_microsat_16xG1 range=chr2:187927363-187928211 5'pad=400 3'pad=400 strand+= repeatMasking=None
CTAGATGCTTAAATACGGCATATGCTGATTTGCTGTAGTGAGTGTGAGCA
GAGACAGCAACAAACAGACACACCTTCTCTCCAGCTTCTT
AGTTCCCTCCTGGGAGACCTCTTCTTCTGGACATCCCCTACAGGGAGCCG
TGGCAAGGGGAAATGTTTGGCCCTGCCCTAGCCCCAGCATCTAGTGT
AGCATGGAAATGTGEECCTAAAGGGGAAAGACAGTATATTAAACAGC
CAAACATTGCCAACACATTAAAAATGATAATGGAAAGAGTGTTTTA
AGAAAAAAATACAGATTAAAAAGGGTCCCAGAAAGAAATAAACACAAA
GAGGCTTAATGCCCTAAAGAAATTAGTTGTTATAGACCAAGCTTAACTTAA
ACACACACACACACACACACACACACACACACACACACACACACAG
GCATCTGCAACACAGTTCACTCAACTTTGTAACACCTTCAAAGCACACAA
AAITCTGATGCTTAAAGCTCTAAAGCATTAGCAATAAAGGATGTT
CTTTAAACATACTCTTTAAACTACTGCTGAGAAGTAAAGGGAAAAAAAG
AAAATATATGCTAATTTTATTATGTTATTAATATATGATGTTGCTTA
AAATACATACCTGAAATAATTAGAGTGAATTGCTAGAAATAAAATAACT
TTTGTCTGTTACACTTCAAAATGTTAAAGTGGATATGTTCTTACTT
CTCTTTTTTGTGAGTAAATAACACAGTTATAATGTTGACTAA
TAATGGGTTATGACATTGGGTTGATTAATTACTCTACAGATATAAG
>hg19_microsat_18xG1 range=chr2:18799496-187995744 5'pad=400 3'pad=400 strand+= repeatMasking=None
CTATTTGGGTTTCTCTTGTGTTTCAAGTATTACAGATGAC
TAGCTTACGTTTATGAAATAGGAGCATGGAGAAAGACTG
CATCCCTCTTCTTGTGAAAGGCTCAGGTTAAATGGGGGTTTAAAGATGAT
TTAGAAATCTGCAATATACCTCCAGAAACATCTCTTAAATCTCTCAT
TTTATCTGCAAGGATGATGTTATTCATGTTAGTGTGATGATGAGTATG
GTATGTTACATATAATACAAACATCAATTTGTAAGTGTGTTTACAA
CTGAGGTTGTTACATATAATTTCTCAAGGAAACATGACACAA
TATTAATGAGTAAATGTTGTTTATGTTGTTGTTGTTGTTGTTGTTGTT
CTGAGGTTGTTACATGAGCTTGTGTTTATGTTGTTGTTGTTGTTGTT
TGTAAATGCTCCATTAATGAGTAAATGTTGTTGTTGTTGTTGTTGTT
TTAAATACACCTGTTGTTAATTTTCTGCAACATATTCTGATGTTGTTA
ACAGATGAAAGACACAGTAAATGACACTTGGCTCTCATAAACATTC
ATTGCTCTACAGGAGTTGTTAATTAATATTTACATGAAATGTTCTGAGT
ATACAGAGGAAACAAAGCCTACTCAATGGTAAATGCTTATTTTAAATGCTT
AGAGTTGTTACATGAGCTTGTGTTTATGTTGTTGTTGTTGTTGTTGTT
TTCTGTTAAACCTGAGTAAATCTGAGGAAATTTTTATGAGATAAG
CATCACACCTTAAAMGGCAACAGGGAGAGTGGCG
>hg19_microsat_17xG1 range=chr2:188000336-188007179 5'pad=400 3'pad=400 strand+= repeatMasking=None
GCATGAGGAAAAATGGTAAATGTTCTACTAGGGCAAGATTTTAAAGCAT
GAAGTTAACTAAATGTTAAATGCAACAGGTTAAAGCATTCTTCTGAGA
AAATAAAAGGTTAGTAACTACAAAGGGCATAGATTAAACAGGTTAAGGCAT
CAAATTCCTAAATTCTCCAGAAGGGAGACTTCTACATAATGTAACAC
ATTTTTATGAGTAAATGTTGCTTAATTCTACTGAGGAGTGA
```

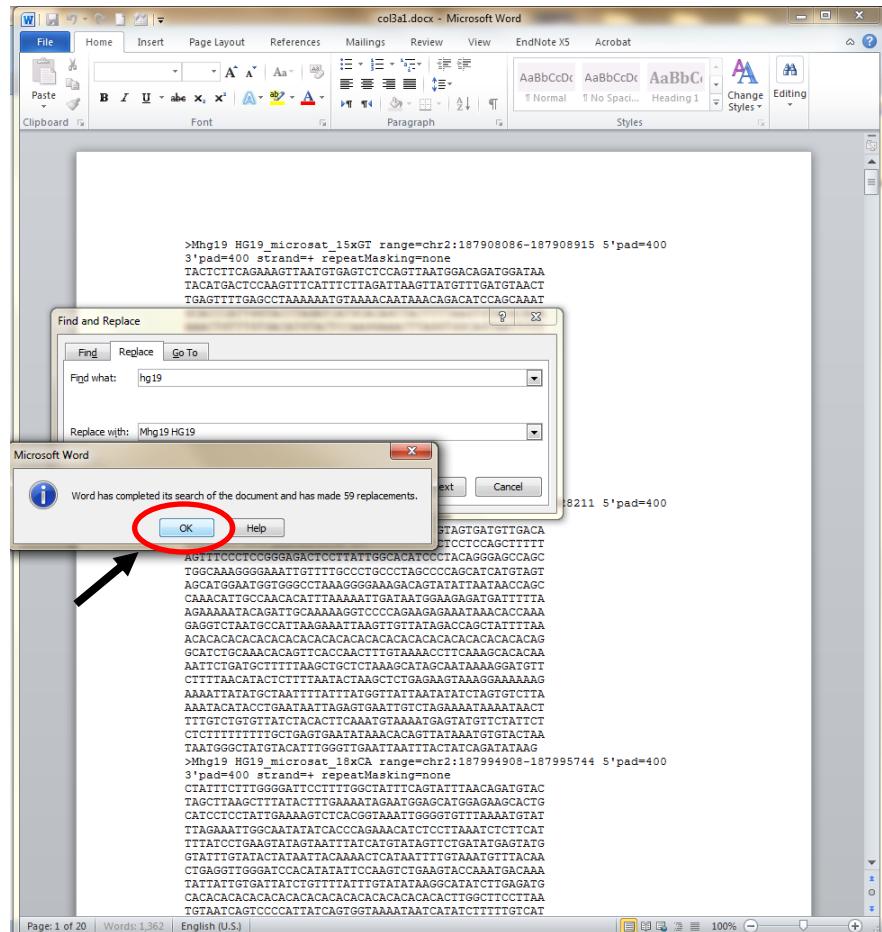
16) Paste the text into a new blank document in Microsoft Word



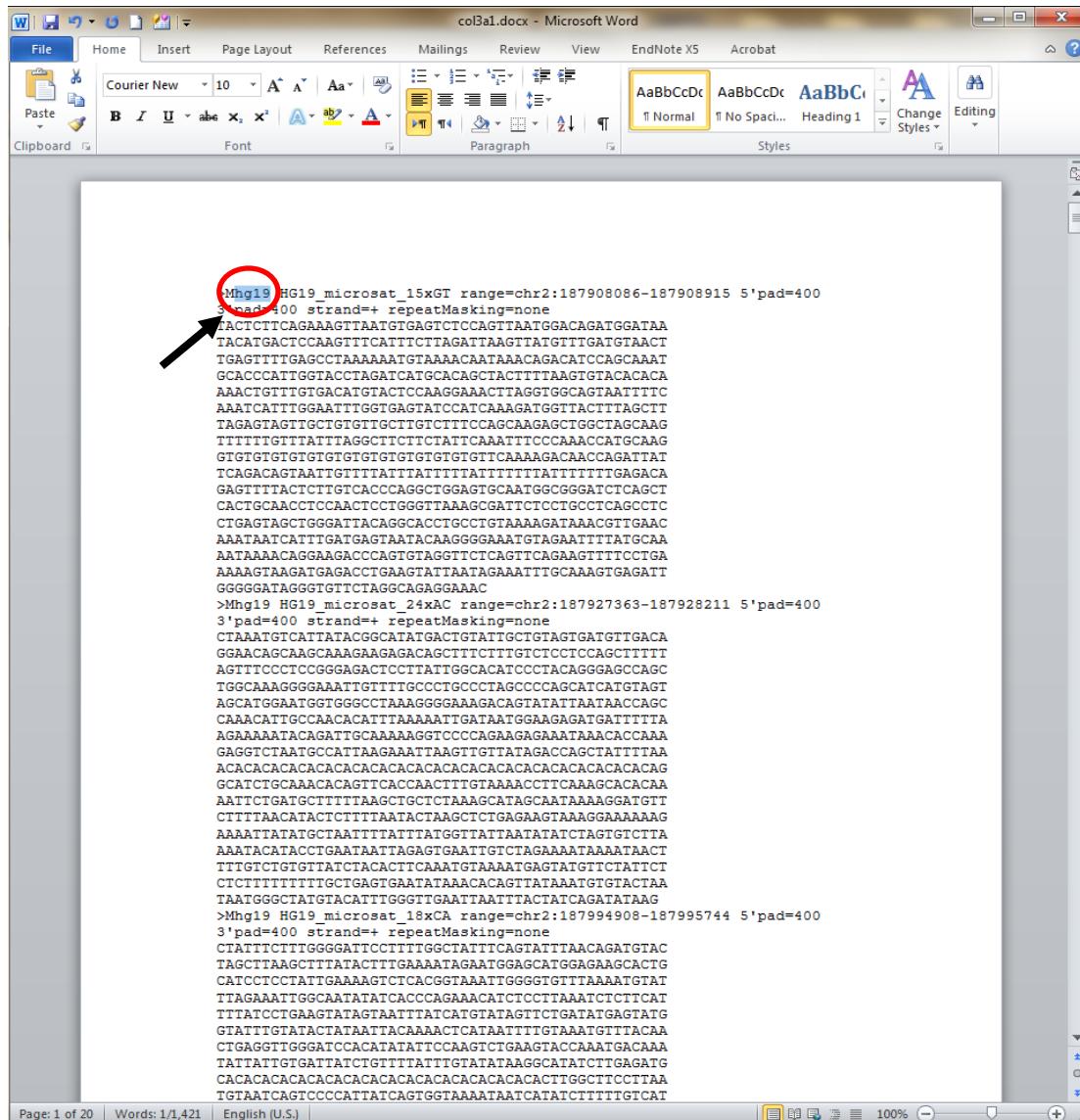
- 17) Press **Ctrl+H**. Click on the “Replace” tab. Type “hg19” next to “Find what:” and type “Mhg19 HG19” next to “Replace with:” (note: it is extremely important to type in the text EXACTLY as written here [including the space between “Mhg19” and “HG19”]). Click “Replace All”



18) Select “OK”

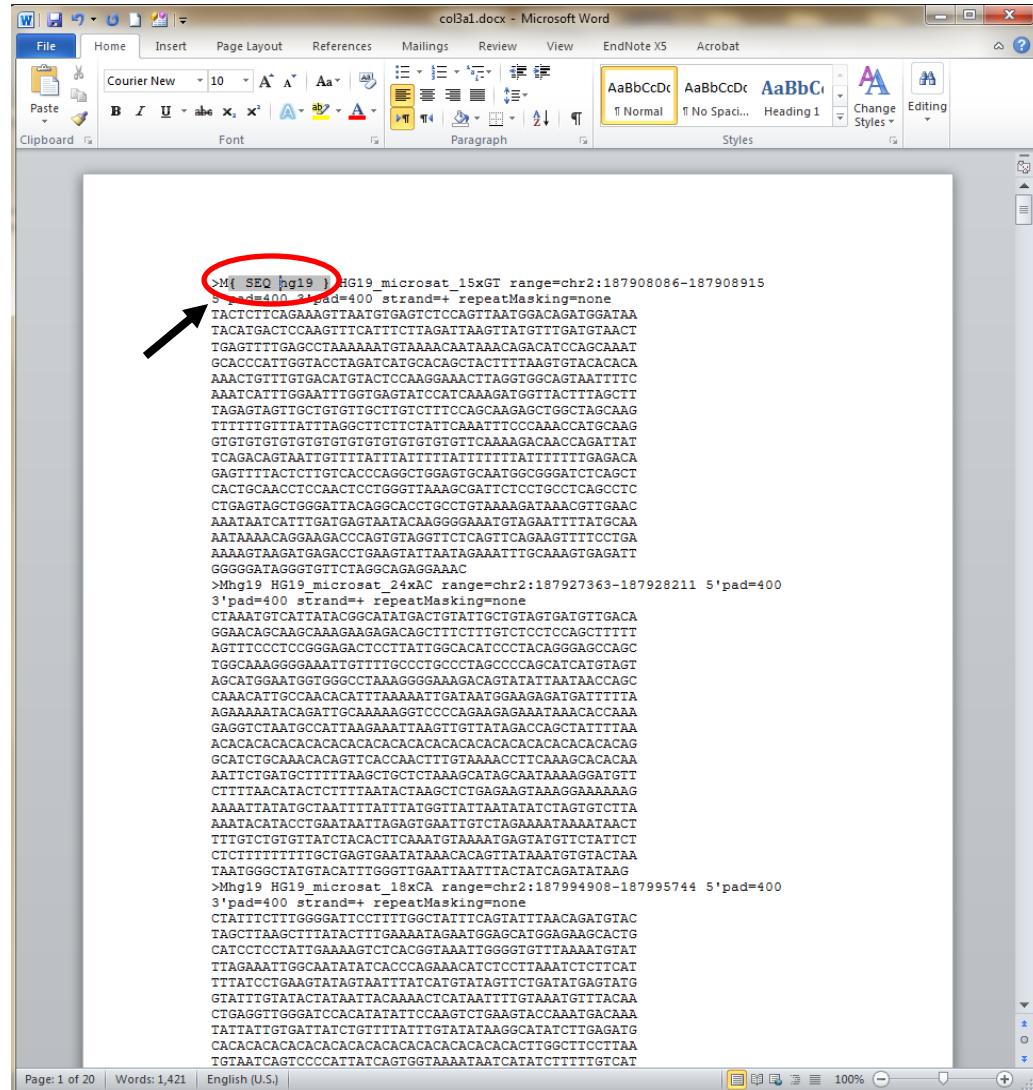


19) Select the first "hg19"



20) Press Ctrl+F9

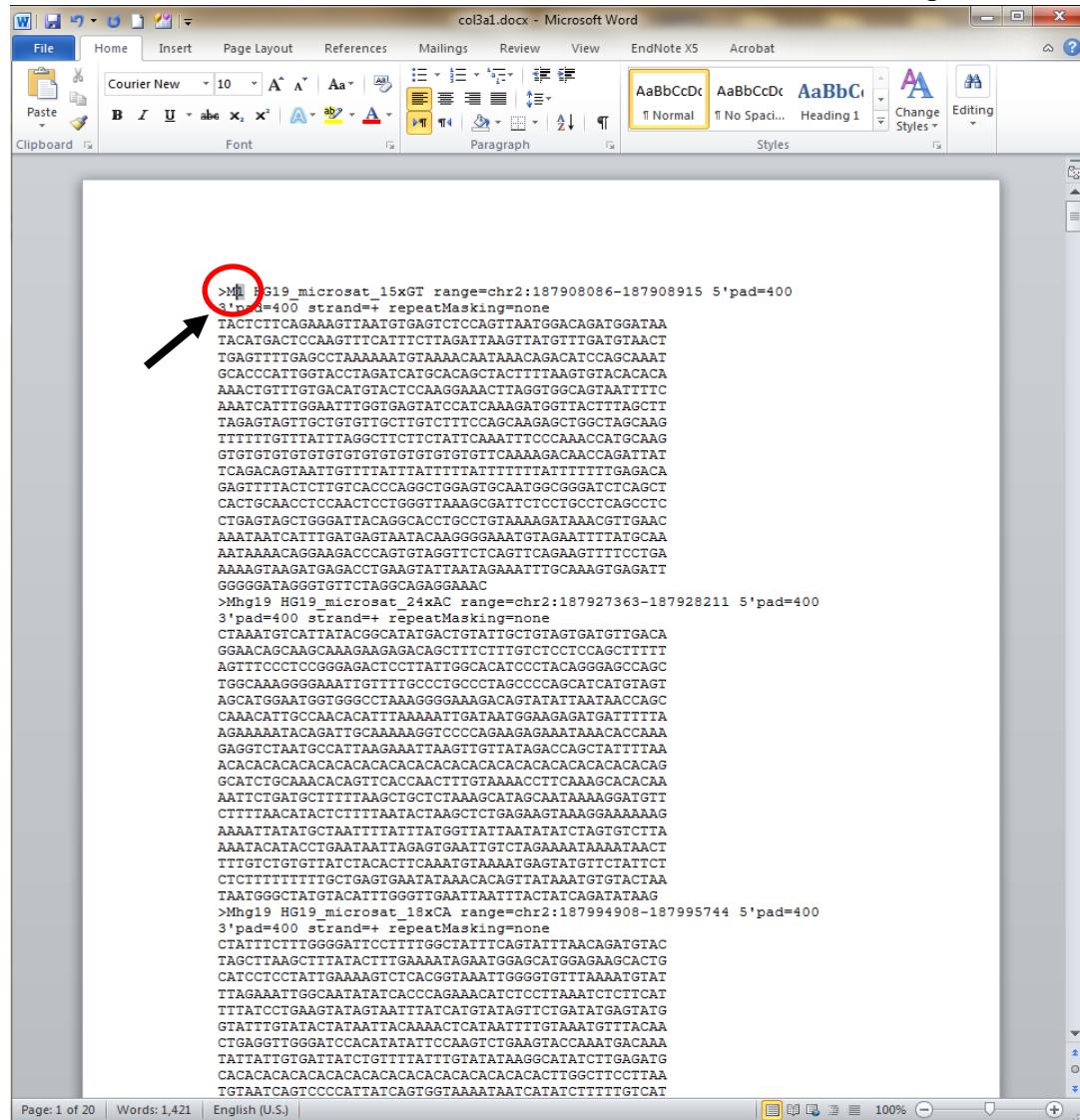
21) Type “SEQ” next to the “hg19” so that it looks like this:



The screenshot shows a Microsoft Word document window titled "co3a1.docx - Microsoft Word". The ribbon menu is visible at the top. A black arrow points from the text "SEQ hg19" to the word "hg19" in the sequence. The sequence itself is a multi-line text block starting with ">M(SEQ hg19) HG19_microsat_15xGT range=chr2:187908086-187908915". The sequence consists of several lines of DNA sequence data, including "5'pad=400 3'pad=400 strand=+" and "repeatMasking=None". The text is formatted in a monospaced font, likely Courier New, and is enclosed in a code block.

```
>M( SEQ hg19 ) HG19_microsat_15xGT range=chr2:187908086-187908915
5'pad=400 3'pad=400 strand=+ repeatMasking=None
>TACATGACTCCAAAGTTCTATTCTTAGATAAAGTTATGGATGTAAC
TGAGTTTGAGCTAAAAAATGTAACAAACATCAGACATCGCAAAT
GCACCCATTGGTACCTAGATCAGACACTTITAAGTGTACACACA
AAACTTTTGAGCATGTACTCAAGGAAACCTAGGTGGCAGTAATTTC
AAATCATTGGAAATTGGTGGATATCCATCAAGATGGTTACTTGTGTT
TAGAGTAGTTGCTGTGTTGCTTCCAGCAAGGCTGGCTAGCAAG
TTTTTGTGTTATTTAGGCTCTCTATTCAAATTCTCCAAACCATGCAAG
GTGTTGTTGTTGTTGTTGTTGTTGTTGCTTCAAGAACAGCAAGATTAT
TCAGACAGTAATTGTTTATTTATTTATTTATTTATGAGACA
GAGTTTACCTGTCACCCAGGCTGGAGTGCATGGCGGAATCTCAGCT
CACTGCAACCTCCAACACTCTGGGTTAAAGCATTCTCCTGCCTCAGCCTC
CTGAGTAGCTGGGATTACAGCACCTGCTGTAAAAGATAACCTGTAAC
AAATAATCATTGATGAGTAATCAAGGGAAATGAGATTATGCAAA
AATAAAACAGGAAGACCCAGTGTAGGTTCTAGTCAAGAAGTTCTGTA
AAAAGTAAGATGAGACCTGAAAGTATTAATAGAAATTGCAAAGTGAGATT
GGGGGATAGGGTGTCTAGGCAGAGGAAC
>Mhg19_HG19_microsat_24xAC range=chr2:187927363-187928211 5'pad=400
3'pad=400 strand=+ repeatMasking=None
>CTAAATGTCATTATACCGCATATGACTGTTGAGTGTGACA
GGAAACAGCAAAGAAAGACAGCTTCTTGTCTCCAGCTTTT
AGTTTCCCTCGGGGAACTCTTATGGCACATCCTCAAGGGGCCAGC
TGGCAAAGGGAAATTGTTTGTGCTTGCCTAGCCCCAGCATCTAGT
AGCATGGAAATGGCCTAAAGGGAAAGCAGTATTTAAACCAAGC
CAAACATTGCCAAACATTAAAGGGTGTGAAAGATGATTTA
AGAAAAATACAGATGCAAAAGGTCGCAAGAGAGAAATAACACAAA
GAGGCTTAATGCTTAAAGAAATTAAAGTGTATAGACAGCTTATTTAA
ACACACACACACACACACACACACACACACACACAG
GCATCTGCAAACACAGTCACCAACTTGTAAAACCTTCAAAGCACACAA
AATTCTGATGTTTTAAGCTGCTCTAAAGCATAGCAATAAAAGGATTT
CTTTAACATACTTTAAATACCTAAAGCTCTGAGAAGTAAAGGAAAAAAG
AAAATTATATGCTAAATTATATGTTTAAATATATCTAGTGTCTTA
AAATACATCTGAAATAATAGACTGAAITGCTAGAAAATAAAACT
TTGCTGTTGAACTTACACTTAAAGTAAATGAGTATGTTCTATCT
CTCTTTTTTTGCTGAGTGAATAACACAGTTAAATGTTACTAA
TAATGGCTATGTACATTGGGTTGAAATTAACTATCAGATATAAG
>Mhg19_HG19_microsat_18xCA range=chr2:187994908-187995744 5'pad=400
3'pad=400 strand=+ repeatMasking=None
>CTATTCCTTGGGATTCCTTGTGGCTATTCACTGATTAAACAGATGTAC
TAGCTTAAGCTTIACTTGTGAAATAGAATGGACATGGGAGAACGACTG
CATCCCTCTTAITGGAAAGCTCAGGTAAATGGGGTGTAAATGTAT
TTGAAAATTGGCAATATACCTCCAGAAACATCTCTTAAACCTCTTAT
TTATCCTGAAAGTATGTAATTATCATGTATAGTCTGATAGTGAAGT
GTTTTGTTATACATTAACACATTAATTGTTGAAATGTTACAA
CTGAGGTTGGGATCCACATATATCCAACTGTAAGTCAACATGACAAA
TATTATTTGTTGATTATCTGTTTATTGTTATATAAGGCATATCTGGAGATG
CACACACACACACACACACACACACACACTTGGCTTCCTAA
TGTAAATCAGTCCCATTATCAGTGGTAAATAATCATATCTTTGTCTA
```

- 22) Press F9 to hide the field codes. You should now see a "1" instead of "hg19". Select the "1" and copy it by pressing Ctrl+C



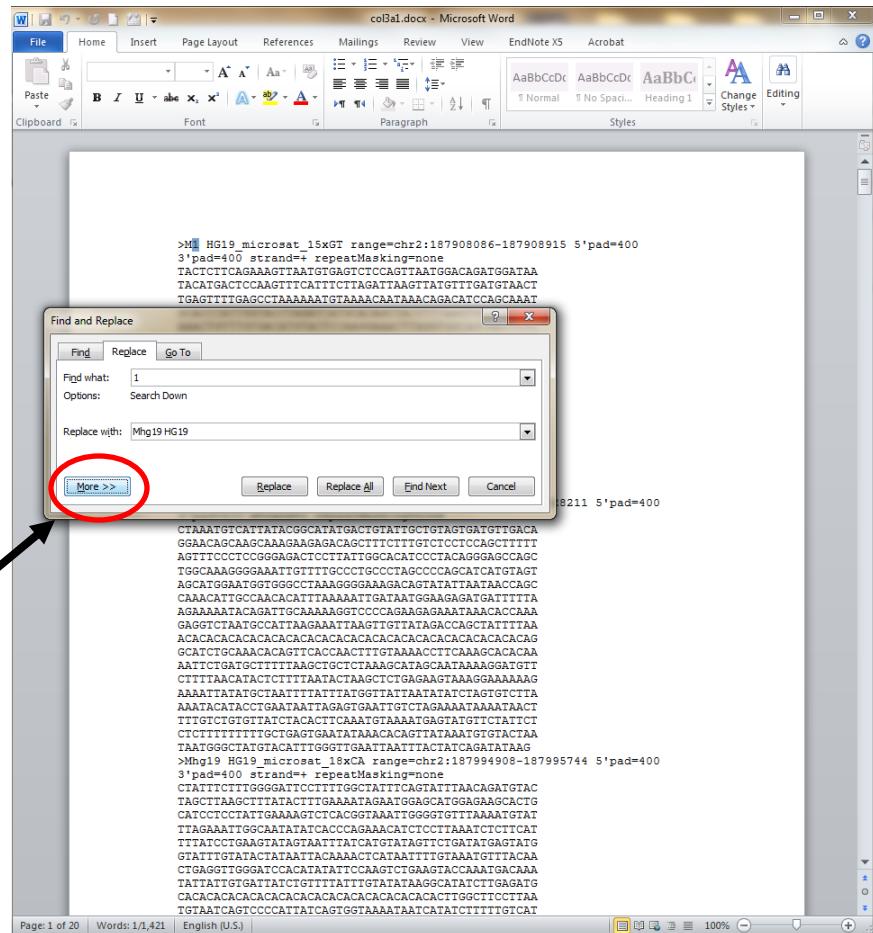
```

>hg19 hg19_microsat_15xGT range=chr2:187908086-187908915 5'pad=400
1'pad=400 strand=+ repeatMasking=None
TACATGACTCCAAGTTAACATGGCTCCAGTTAATGGACAGATGGATAA
TGAGTTTGAGCCCTAAAAAAATGAAAATAAACAGACATCAGCAAAAT
GCACCCATTGGTACCTAGATCATGCACAGCTACTTTAACGTGACACACA
AAACTGTTTGTGACATGACTCCAAAGGAACATTAGTTGGCAGTAATTTC
AAATCATTGGAAATTGGTAGTATCCATCAAAGATGGTACTTTAGCTT
TAGAGTAGTTGCTGTGCTTCTTCCAGCAAGAGCTGGCTAGGAAG
TTTTTGTATTAGGCTCTCTTCAATTCAAAATTCCCAACATGCAAG
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TCAGACAGTAATTGTTTATTATTATTATTATTATTATTGAGACA
GAGTTTTAACCTTGTACCCAGGCTGGAGTGCATGGCGGAACTCAGCT
CACTGCAACCTCCAAACTCTGGTTAAACCGATTCTCTGCCCTAGCCTC
CTGAGTAGCTGGATTACAGGGACCTGCTGTAAAAGATAAACGTTGAC
AAATAATCATTTGATGAGTAATACAAGGGGAAATGTAGAATTTTATGCAA
AATAAACAGGAAGACCCAGTGTAGGTTCTCAGTCAGAAGTTCTGAA
AAAAGTAAGATGAGACCTGAAGTATAATAGAAATTTCAGAAAGTGAGATT
GGGGATAGGGTGTCTAGGCAGAGGAAC
>hg19 hg19_microsat_24xAC range=chr2:187927363-187928211 5'pad=400
1'pad=400 strand=+ repeatMasking=None
CTAAATGTCATTATACGGCATATGACTGTATTGCTGTAGTGATGTTGACA
GGACAGCGAACAAAGAGACAGCCTTCTTGTCTCCACGCTTCTT
AGTTTCCCTCCGGAGACTCTTATGGCACATCCTACAGGGAGCCAGC
TGGCAAAGGGAAATTGTTGCTGTGCTGGCTTACGCCCAAGCATGTAGT
AGCATGGAAATGTTGGCTAAAGGGAAAGACAGTATATTAAACACAGC
CAAACATTGCAACACATTAAAAAATGATAATGGAAAGAGTGTATTAA
AGAAAAATAACAGATTGCAAAAGTCCCCAGAGAAATAACACCAAA
GAGGTCTAATGCCATTAAAGAAATTAAAGTTGTTATAGACCACTTAA
ACACACACACACACACACACACACACACACACACACACACAG
GCATCTGCAAACACAGTCACCAACTTGTAAAACCTTCAACGACACAA
AATTCTGATGCTTTAAAGCTCTAAAGCATAGCAATAAAAGGATGTT
CTTTAACATACTTTAAACTAAAGCTCTGAGAATATAACACAGTTAAATGTC
AAAATTATATGCTAATTTTATGTTTAAATATATCTAGTGTCTTA
AAATACATACTGAAATAATTAGAGTGAATTGCTGAAATAAAATAACT
TTTGTGTGTATTCTACACTTCAAAATGAAATTAGTATGTCTATTCT
CTCTTTTTTTGCTGAGTGAATATAACACAGTTAAATGTCAGTAA
TAATGGCTATGTCACATTGGGTTGAATTAATTACTATCAGATATAAG
>hg19 hg19_microsat_18xCA range=chr2:187994908-187995744 5'pad=400
1'pad=400 strand=+ repeatMasking=None
CTATTCTTGGGGATTCTTITGGCTATTTCAGTATTTAACAGATGTAC
TAGCTTAAGCTTATACCTTGAAGATGAAATGAGGAGCATGGAGAACACTG
CATCCCTCTATTGAAAAGCTCACCGTAAATTGGGGTGTAAAAATGTAT
TTAGAAATGGCAATATACCTCCAGAAACATCTCTTAAATCTCTTCTAT
TTTATCTGAAATGATGTTATCATGTATGTTCTGATATGAGTATG
GTATTGTTATACATAATTACAAACACTATAATTGTTAAATGTTACAA
CTGAGGTGGGATCCACATATTCCAAAGCTGAGTACCAAAATGACAAA
TATATTGTTGATTCTGTTTATTGTATATAAGGSCATATCTGAGATG
CACACACACACACACACACACACACACACACACTTGGCTTCTAA
TGTATCAGTCCCCATTATCAGTGGTAAAAATAATCATATCTTTGTCAT

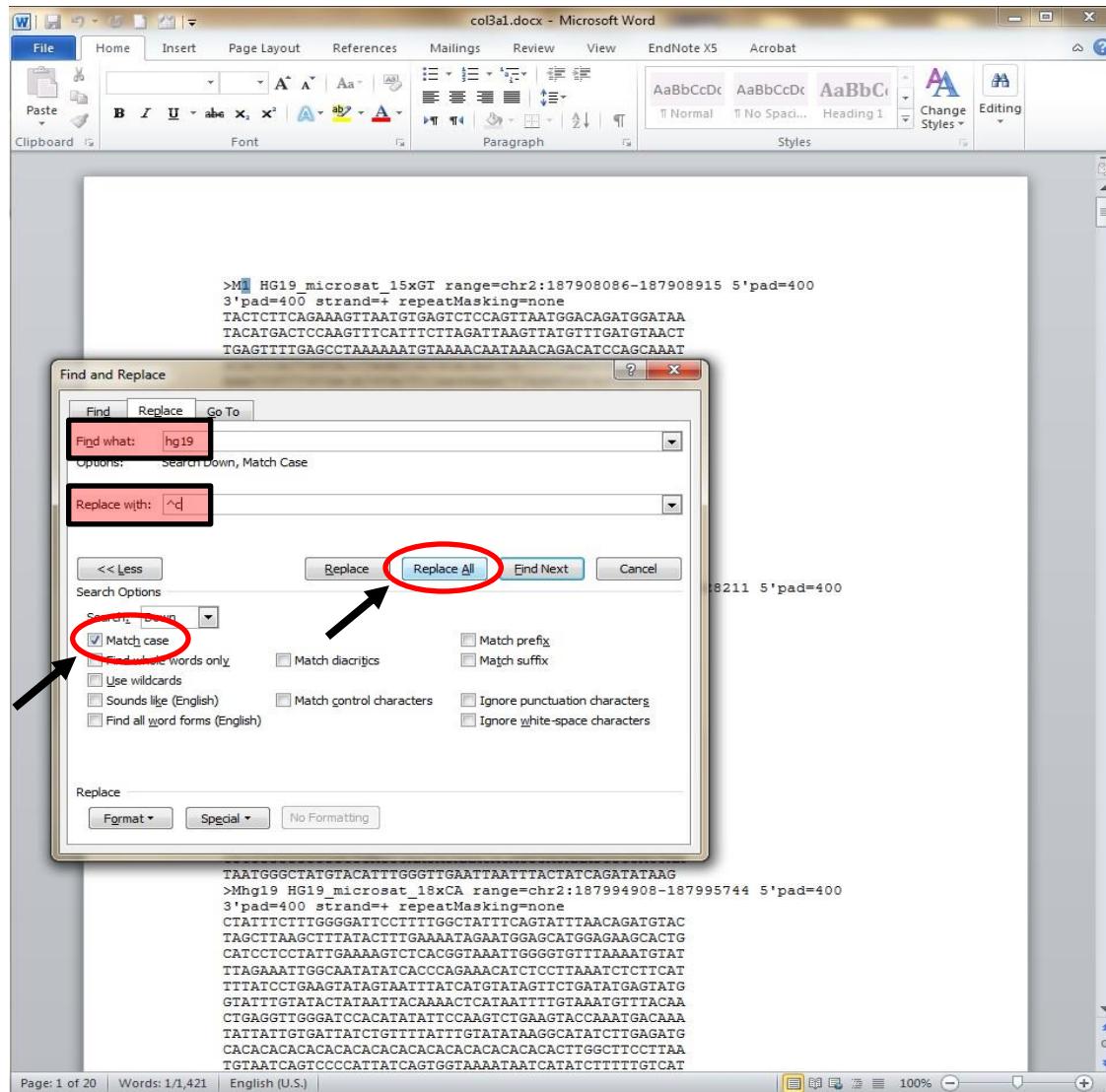
```

Page: 1 of 20 | Words: 1,421 | English (U.S.) | 100% | |

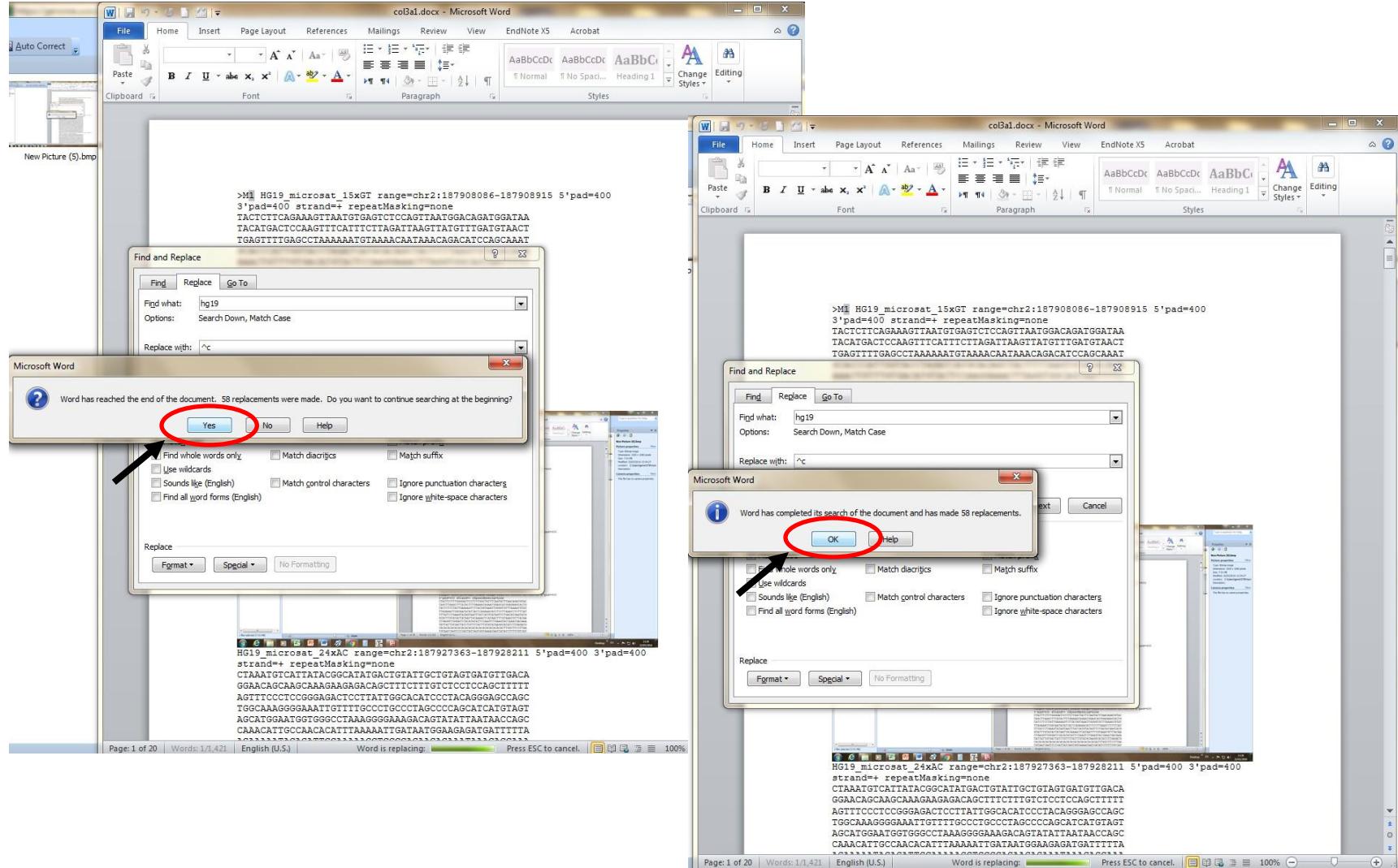
23) Press Ctrl+H. Click on “More>>”



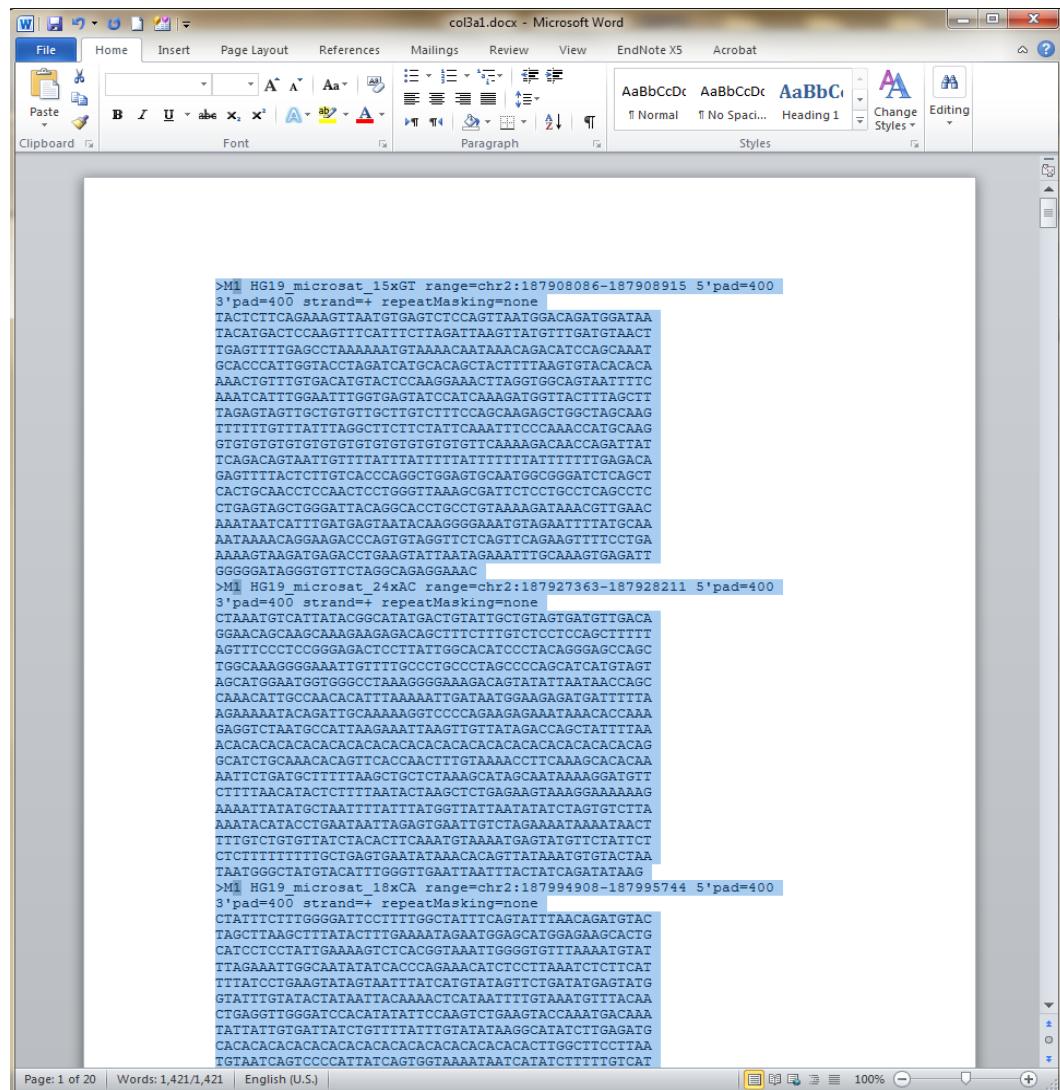
- 24) Select “Match case”. Type “hg19” next to “Find what:”. Type “^c” next to “Replace with:” (note: ^c is the code for what’s on the clipboard). Click “Replace All”



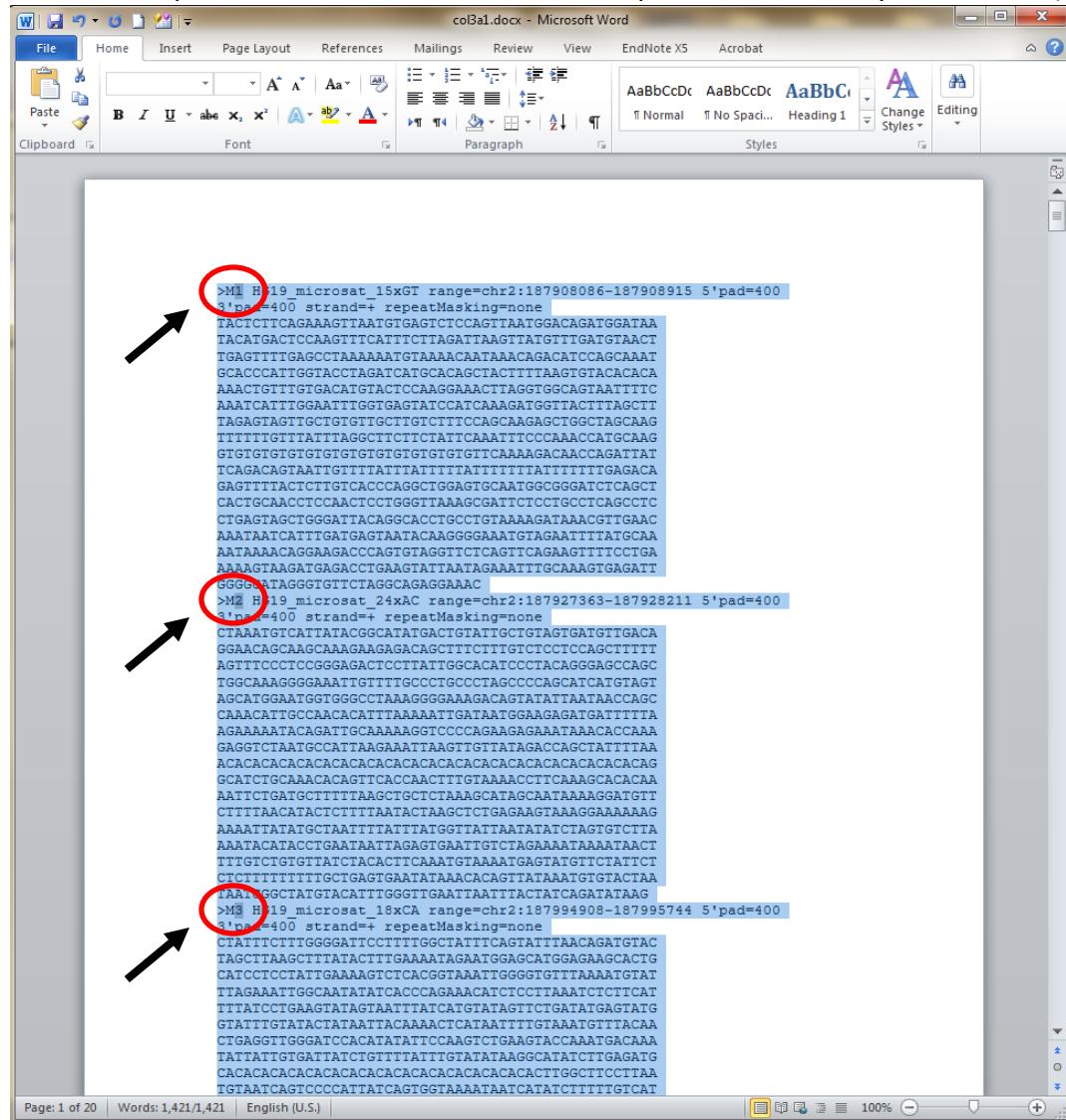
25) Click "Yes" then click "OK"



26) Press Ctrl+A to select the entire document



27) Press F9 to update fields. You should now see sequential names for your markers (">M1" ; ">M2" ; ">M3" ; etc...

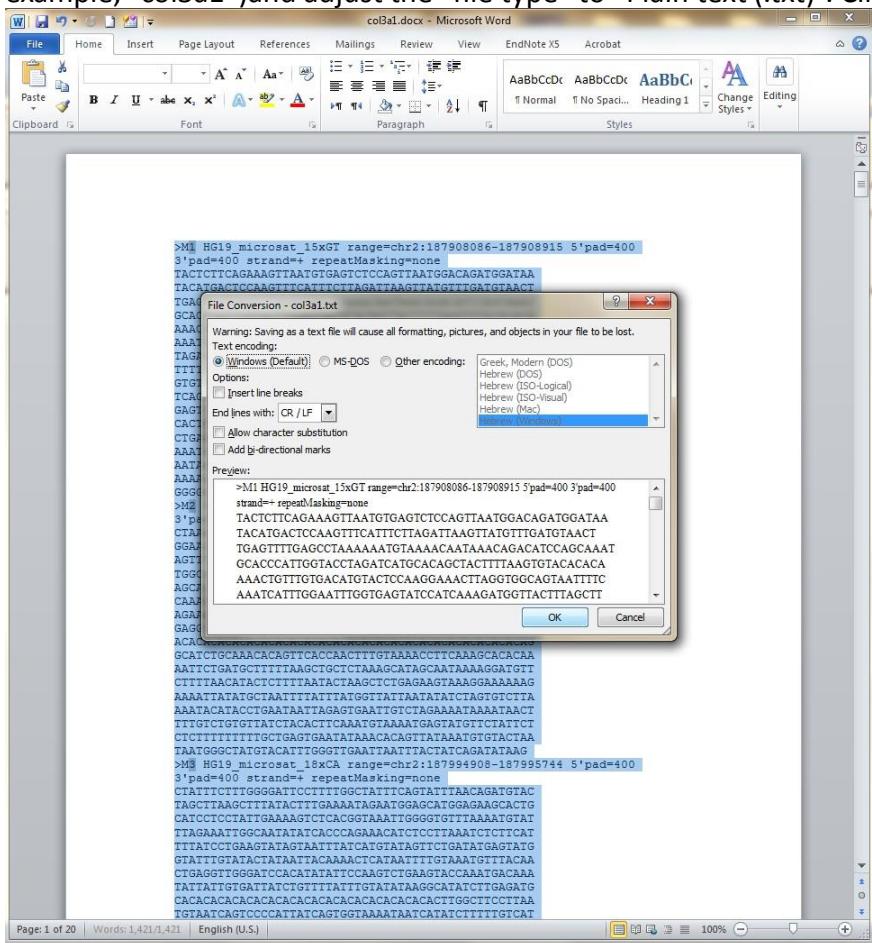


The screenshot shows a Microsoft Word document window titled "coB1.docx - Microsoft Word". The ribbon menu is visible at the top. The main content area contains three genomic sequence blocks, each preceded by a marker name and parameters:

- >M1 H_19_microsat_15xGT range=chr2:187908086-187908915 5'pad=400
3'pa=400 strand=+ repeatMasking=none
- >M2 H_19_microsat_24xAC range=chr2:187927363-187928211 5'pad=400
3'pa=400 strand=+ repeatMasking=none
- >M3 H_19_microsat_18xCA range=chr2:187994908-187995744 5'pad=400
3'pa=400 strand=+ repeatMasking=none

Each sequence block is highlighted with a red circle. Three black arrows point from the left towards these red circles, indicating the specific markers being referred to in the instruction.

- 28) In the Word menu, select “File” > “Save as”. Choose a path to save your file on your computer. Type in a file name (for example, “col3a1”) and adjust the “file type” to “Plain text (.txt)”. Click “Save”. Click “OK”



- 29) Access the Hemi-NeSTR website. In step 3 “Paste your sequence(s)...” click “Choose File”. Navigate to the file you just saved (for example col3a1.txt) and click “Open.”
- 30) You may now click “Submit Job” if all other design parameters are suitable for your project design.