Experiment - 1

Name: **Suchismita Tripathy**
ID: **2019A7PS0554P**
Gene Name: **HUS1 Checkpoint Clamp Component (HUS1) transcript variant X1**
Organism: **Theropithecus gelada**
Accession Number: **XM_025380797.1**

I. **Aim:** To retrieve the nucleotide sequence of Theropithecus gelada HUS1 Checkpoint Clamp Component (HUS1) transcript variant X1 gene in Fasta and Genbank format using NCBI

II. **Materials Required:** NCBI, Fasta and GenBank format

III. **Procedure:** Steps:

1. Go to the NCBI homepage.
2. Select "Nucleotide".
3. Enter "Accession No" (XM_025380797.1) in the search field and click the "Search" button.
4. Select specific format.
5. Interpret the results

IV. Observations:

A. FASTA

```
>XM_025380797.1 PREDICTED: Theropithecus gelada HUS1 checkpoint clamp component (HUS1),
transcript variant X1, mRNA
AACCATGTGCGCGCCGCCGTGTTTCCGGGCGGGGACACTCAGGGCGCGACACTCATCTGTTACCCACAGA
GGCCGTCGCGGCTGCGGCAGGCGCGGCCATGAGGTTTCGGGCCAAGATCGTGGACGGGGCCTGTCTGAAC
CACTTCACACGAATCAGTAACATGATAGCCAAGCTTGCCAAAACCTGCACCCTCCGCATCAGCCCTGATA
AGCTTAACTTCATCCTTTGTGACAAGCTGGCTAATGGAGGGGTGAGCATGTGGTGTGAGCTGGAACAGGA
GAACTTCTTCAACGAATATCAAATGGAGGGTGTCTCTGCAGAAAACAATGAGATTTATTTAGAGCTAACA
TCGGAAAACTTATCTCGAGCCTTGAAAACTGCCCAGAATGCCAGAGCCTTGAAAATCAAACTGACTAATA
AACACTTTCCCTGCCTCACAGTCTCCGTGGAGCTGTTATCTATGTCAAGCAGTAGCCGCATTGTGACACA
TGACATCCCTATAAAGGTGATTCCTAGGAAATTGTGGAAGGACTTGCAAGAACCGGTGGTCCCAGATCCT
GATGTTAGTATTTATTTACCAGTCTTGAAGACTATGAAGAGTGTTGTGGAAAAAATGAAAAACATCAGCA
ATCACCTTGTTATTGAAGCAAACCTAGATGGAGAATTGAATTTGAAAATAGAAACTGAATTAGTATGTGT
TACAACTCATTTTAAAGATCTTGGAAATCCTCCTTTAGCCTCTGAAAGCACCCATCAAGACAGAAACATG
GAACACATGGCTGAAGTGCACATAGATATTAGGAAGCTCCTACAGTTTCTTGCTGGACAACAAGTAAATC
CCACAAAGGCCTTATGCAATATTGTGAATAACAAGATGGTGCATTTTGATCTGCTTCACGAAGACGTCTC
CCTTCAGTATTTCATCCCTGCGCTGTCCTAGCACCCTGTTGCTGGAGTTGGCATGCGGAGACTTTGTCAG
GATGGGAGAGGCCGTAGGCGTTGTGTTCTGATCACTGGTCTGTGTCCTCACAGCACCGCACATCGACACA
CTGTACTTATTTGTCCCTCTCTAACATTTTAACTAAAAGTTGACTTAATGGCACACAGTTGGATAAACAT
ATCACTTCATGTTGCTCATGTCTGTTTTGTTTTGTTTTTAAGACATTGAAAAGAAAAGCTAGAATTTATT
TATTCAGACTTTAAAGAACAATTTCTCATTGACGTTGTGAAAATCCTCATGTATTTAGACTTGGTGTAGT
AGCCAGAATTCGTGCAGCTGTTGCCTGGGAGCTGGGTACTTTCCCTCCGGGCAGAGGCTCTAGCTCAGCA
CGGCCTGTAGCGCACAGTCAGTCTTGCATTTCAGTGTGTTCACCCTGCTTGCTCCTGCCCCTTGGAGCCC
AGTGACAGAAAGTATAGCCTCTGTCACCCCGCTGCCACTGCCTTGGTTACTCAGAGCGCTGTGGGGTGTC
ACAGCTGCAGCATTTGGGGTCTCTCTCTCTCTTGCTGAGTACTCAAGCCCACCTGAGTCCACTCCCCTCT
TGATGCCTGGAGAGCTGGCCCAGCCAACACAGCTTTTCGCTGGGAGCTCCTTCTGCCATTCCAATTAGTT
TCTTCCTGGGGCCAGTTTTGGGTTTAGGTTGTAATTCCTTATATTTCTTTCTTCCACAGTGCATCGGATC
TGTCATTCTGGAAAGAAGACCCTTCTATTTAGAGTAGAAACAAATGAAACTTCTAAGGTATCATCTGTGT
TAGGTGATGAGACCATATTTCTTTGATGTTTCTGAACATCAAAGCTGATTCAGTACTGGTAGATGTGCTC
ATTCTCCCTGAAACATACCTATCATATTTCCTATTATAATTCTATCTCATTGTCCTGTGGAGGTGGACAT
GATCAACAATATCTTTTATTTTCTTGTTTTGTTTTGTTTGAGACAAGGTCTCACTCTGTCACCCAGACTG
GAGTGAAGGGCAACAATCATGGCTCACCGCCTTGACCTCCTTGGCTCAAGGCATCCTCCCACCTCAGCCT
CCTGACTACCTGGGACTACTGGTGTGCGCCACCACATCCAGCTAGTTTTTAATTTTTCATAGAGACAGAG
TCCCACTGTGTTGCCTAGGCTGGTCTTGAACTCCTGGGCTCAAGCCACCCACCCACCTGGGCCTCCCAAA
GTGCTGGGATTACAGGCATTAGCCATCACACCCATCCGTAAACGTTATCTTAATGTCACATTACAAAACT
GAGGCCTAAGTTGCTTAGGACAAGATGAAGAAATCGAAGACTAGCTAACATGAAAATTTATATTTTGGCT
TTTTCATGTTTTTTTGGTAAAACCAGTGTATTTGAATAGTTCTTTTGATGTTTCATAATGGTTTTTTGTTT
GTTTGCTTGGTTCAGTTTTTTTTTTTCCTTGAGACAGAGTCTTGCTCTGTCGCCAGGCCAGAGTGCCAGT
GTCACGATCTTGGCTCACTGCAACCTCCACCTCCCAAGTTCAAGCGATTCTCCTGCCTCAGCCCCCCGAG
TAGCTGGGACTATAGGCACGTGCCACCACACCCAGCTAATTTTTGTATTTTTAGTAGAGATGGCATTTCA
CTATGTTGGCCAGGATGGTCTCGATCTCTTGACTTTGTGATCCACCTGCCTTGGCCTCTCAAAAGTGCTG
GGATTACAGGCGTGAGCCACCATGCCTGGCCTGTGTTTAATAATGTTTAAATAGGGTGGAATATTTTGTT
AAATTAACATTTTAAAATTAGAAGACGCCATTTTAATTTTTAAACCCTTTCTCCTCGTTGTAACAAAATT
AATTCCAGCTGTAGTGAGAAAACTTAAAAATCATGATACAAAATGAAACAATATCTGAAAGTAGTTTTAT
AAAACTGAAATTATTGTTAAAGAGAATGGTATTAGTGACTTAACCATTTGCTCTATATGATGTTTATTAT
CAAATACACATAATTTTGAAGATTTTAATGAATGGCTTAAGATTTTATCTTTGTGTAGAATGTGGCTAAA
GAAACCTTAGTTGAGATTCAA
```

B. GenBank

```
LOCUS       XM_025380797          3031 bp    mRNA    linear   PRI 29-JUN-2018
DEFINITION  PREDICTED: Theropithecus gelada HUS1 checkpoint clamp component
            (HUS1), transcript variant X1, mRNA.
ACCESSION   XM_025380797
VERSION     XM_025380797.1
DBLINK      BioProject: PRJNA477372
KEYWORDS    RefSeq.
SOURCE      Theropithecus gelada (gelada)
  ORGANISM  Theropithecus gelada
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Cercopithecidae; Cercopithecinae; Theropithecus.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NC_037670.1) annotated using gene prediction method: Gnomon,
            supported by mRNA and EST evidence.
            Also see:
                Documentation of NCBI's Annotation Process

            ##Genome-Annotation-Data-START##
            Annotation Provider          :: NCBI
            Annotation Status            :: Full annotation
            Annotation Name              :: Theropithecus gelada Annotation
                                            Release 100
            Annotation Version           :: 100
            Annotation Pipeline          :: NCBI eukaryotic genome annotation
                                            pipeline
            Annotation Software Version :: 8.1
            Annotation Method            :: Best-placed RefSeq; Gnomon
            Features Annotated           :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..3031
                     /organism="Theropithecus gelada"
                     /mol_type="mRNA"
                     /isolate="Dixy"
                     /db_xref="taxon:9565"
                     /chromosome="3"
                     /sex="female"
                     /tissue_type="whole blood"
                     /dev_stage="adult"
                     /country="Ethiopia: Simien Mountains"
                     /collection_date="Nov-2017"
                     /collected_by="Simien Mountains Gelada Research Project"
     gene            1..3031
                     /gene="HUS1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Gnomon. Supporting evidence
                     includes similarity to: 8 mRNAs, 85 ESTs, 3 Proteins, and
                     42% coverage of the annotated genomic feature by RNAseq
                     alignments"
                     /db_xref="GeneID:112621393"
     CDS             99..941
                     /gene="HUS1"
                     /codon_start=1
                     /product="checkpoint protein HUS1 isoform X1"
                     /protein_id="XP_025236582.1"
                     /db_xref="GeneID:112621393"
                     /translation="MRFRAKIVDGACLNHFTRISNMIAKLAKTCTLRISPDKLNFILC
                     DKLANGGVSMWCELEQENFFNEYQMEGVSAENNEIYLELTSENLSRALKTAQNARALK
                     IKLTNKHFPCLTVSVELLSMSSSSRIVTHDIPIKVIPRKLWKDLQEPVVPDPDVSIYL
                     PVLKTMKSVVEKMKNISNHLVIEANLDGELNLKIETELVCVTTHFKDLGNPPLASEST
                     HQDRNMEHMAEVHIDIRKLLQFLAGQQVNPTKALCNIVNNKMVHFDLLHEDVSLQYFI
                     PALS"
```

```
ORIGIN
        1 aaccatgtgc gcgccgccgt gtttccgggc ggggacactc agggcgcgac actcatctgt
       61 tacccacaga ggccgtcgcg gctgcggcag gcgcggccat gaggtttcgg gccaagatcg
      121 tggacggggc ctgtctgaac cacttcacac gaatcagtaa catgatagcc aagcttgcca
      181 aaacctgcac cctccgcatc agccctgata agcttaactt catcctttgt gacaagctgg
      241 ctaatggagg ggtgagcatg tggtgtgagc tggaacagga gaacttcttc aacgaatatc
      301 aaatggaggg tgtctctgca gaaaacaatg agatttattt agagctaaca tcggaaaact
      361 tatctcgagc cttgaaaact gcccagaatg ccagagcctt gaaaatcaaa ctgactaata
      421 aacactttcc ctgcctcaca gtctccgtgg agctgttatc tatgtcaagc agtagccgca
      481 ttgtgacaca tgacatccct ataaaggtga ttcctaggaa attgtggaag gacttgcaag
      541 aaccggtggt cccagatcct gatgttagta tttatttacc agtcttgaag actatgaaga
      601 gtgttgtgga aaaaatgaaa aacatcagca atcaccttgt tattgaagca aacctagatg
      661 gagaattgaa tttgaaaata gaaactgaat tagtatgtgt tacaactcat tttaaagatc
      721 ttggaaatcc tcctttagcc tctgaaagca cccatcaaga cagaaacatg gaacacatgg
      781 ctgaagtgca catagatatt aggaagctcc tacagtttct tgctggacaa caagtaaatc
      841 ccacaaaggc cttatgcaat attgtgaata acaagatggt gcattttgat ctgcttcacg
      901 aagacgtctc ccttcagtat ttcatccctg cgctgtccta gcaccctgtt gctggagttg
      961 gcatgcggag actttgtcag gatgggagag gccgtaggcg ttgtgttctg atcactggtc
     1021 tgtgtcctca cagcaccgca catcgacaca ctgtacttat ttgtccctct ctaacatttt
     1081 aactaaaagt tgacttaatg gcacacagtt ggataaacat atcacttcat gttgctcatg
     1141 tctgttttgt tttgttttta agacattgaa aagaaaagct agaatttatt tattcagact
     1201 ttaaagaaca atttctcatt gacgttgtga aaatcctcat gtatttagac ttggtgtagt
     1261 agccagaatt cgtgcagctg ttgcctggga gctgggtact ttccctccgg gcagaggctc
     1321 tagctcagca cggcctgtag cgcacagtca gtcttgcatt tcagtgtgtt caccctgctt
     1381 gctcctgccc cttggagccc agtgacagaa agtatagcct ctgtcacccc gctgccactg
     1441 ccttggttac tcagagcgct gtggggtgtc acagctgcag catttggggt ctctctctct
     1501 cttgctgagt actcaagccc acctgagtcc actcccctct tgatgcctgg agagctggcc
     1561 cagccaacac agcttttcgc tgggagctcc ttctgccatt ccaattagtt tcttcctggg
     1621 gccagttttg ggtttaggtt gtaattcctt atatttcttt cttccacagt gcatcggatc
     1681 tgtcattctg gaaagaagac ccttctattt agagtagaaa caaatgaaac ttctaaggta
     1741 tcatctgtgt taggtgatga gaccatattt ctttgatgtt tctgaacatc aaagctgatt
     1801 cagtactggt agatgtgctc attctccctg aaacatacct atcatatttc ctattataat
     1861 tctatctcat tgtcctgtgg aggtggacat gatcaacaat atcttttatt ttcttgtttt
     1921 gttttgtttg agacaaggtc tcactctgtc acccagactg gagtgaaggg caacaatcat
     1981 ggctcaccgc cttgacctcc ttggctcaag gcatcctccc acctcagcct cctgactacc
     2041 tgggactact ggtgtgcgcc accacatcca gctagttttt aatttttcat agagacagag
     2101 tcccactgtg ttgcctaggc tggtcttgaa ctcctgggct caagccaccc acccacctgg
     2161 gcctcccaaa gtgctgggat tacaggcatt agccatcaca cccatccgta aacgttatct
     2221 taatgtcaca ttacaaaact gaggcctaag ttgcttagga caagatgaag aaatcgaaga
     2281 ctagctaaca tgaaaattta tattttggct ttttcatgtt ttttggtaaa accagtgtat
     2341 ttgaatagtt cttttgatgt ttcataatgg ttttttgttt gtttgcttgg ttcagttttt
     2401 tttttttcctt gagacagagt cttgctctgt cgccaggcca gagtgccagt gtcacgatct
     2461 tggctcactg caacctccac ctcccaagtt caagcgattc tcctgcctca gcccccgag
     2521 tagctgggac tataggcacg tgccaccaca cccagctaat ttttgtattt ttagtagaga
     2581 tggcatttca ctatgttggc caggatggtc tcgatctctt gactttgtga tccacctgcc
     2641 ttggcctctc aaaagtgctg ggattacagg cgtgagccac catgcctggc ctgtgtttaa
     2701 taatgtttaa atagggtgga atattttgtt aaattaacat tttaaaatta gaagacgcca
     2761 ttttaatttt taaacccttt ctcctcgttg taacaaaatt aattccagct gtagtgagaa
     2821 aacttaaaaa tcatgataca aaatgaaaca atatctgaaa gtagttttat aaaactgaaa
     2881 ttattgttaa agagaatggt attagtgact taaccatttg ctctatatga tgtttattat
     2941 caaatacaca taattttgaa gattttaatg aatggcttaa gattttatct ttgtgtagaa
     3001 tgtggctaaa gaaaccttag ttgagattca a
//
```

V. Inferences:

A. Summary: **The protein encoded by this gene is a component of an evolutionarily conserved, genotoxin-activated checkpoint complex that is involved in the cell cycle arrest in response to DNA damage. This protein forms a heterotrimeric complex with checkpoint proteins RAD9 and RAD1. In response to DNA damage, the trimeric complex interacts with another protein complex consisting of checkpoint protein RAD17 and four small subunits of the replication factor C (RFC), which loads the combined complex onto the chromatin. The DNA damage induced chromatin binding has been shown to depend**

on the activation of the checkpoint kinase ATM and is thought to be an early checkpoint signaling event. Alternative splicing results in multiple transcript variants.

B. Location: **chromosome: 3**

Sequence : NC_037670.1 -> 65,152,075 ..65,169,087

C. Exon Count and its Location : **8 for transcript variant x1**

1. **(1 ..150)**
2. **(151 ..278)**
3. **(279 .. 455)**
4. **(456 ..563)**
5. **(564 ..638)**
6. **(639 ..738)**
7. **(739 ..858)**
8. **(859 ..3031)**

D. Base Composition Size: **1 ..3031 : 3031 nt**

E. Coding Sequence Length: **99 ..941 : 843 nt**