# README

SURYA TAMRAPARNI (A53254448)

AMARNATH RAJU (A53079315)

To Run

Package contains StepMiner.py and README.docx.

StepMiner takes a prefix(example : “cse291-test”), which must be a string, and writes prefix-thr.txt, prefix-info.txt, prefix-bv.txt in the same directory. The directory must have write permissions. The input files must be placed in the same directory as the python file.

The following command can be used to run the python script.

* python StepMiner.py "prefix"

**O(n)** Algorithm for Fitting a StepMiner:

*Pseudocode(List of Values of size n):*

1. Sort the data in ascending order, call it l
2. Calculate sstot.
3. Consider 2 sets. Left and Right. Start with 0 (call it n1) elements in Left and n elements in the right set (call it n2).
4. Calculate the means m1 for left set (i.e. 0) and m2 for right set (i.e. mean of all elements , m).
5. Initialize min\_sse = sstot
6. For each element x in L:

* Put element x in Left\_set, recalculate m1. m1 = ((m1\* n1) + x)/(n1+1)
* n1 = n1+1
* Remove x from the right\_set, recalculate m2. m2 = ((m2 \* n2) - x)/(n2 - 1)
* n2 = n2 -1
* Calculate SSR. ssr = (n1 \* ((m1 - m)\*\*2)) + (n2 \* ((m2 - m)\*\*2))
* Calculate SSE. sse = (sstot - ssr)
* If min\_sse > sse:
* min\_sse = sse
* min\_m1 = m1
* min\_m2 = m2

1. return (min\_m1 + min\_m2)/2

**Tissue : Kidney (renal) cell tissue:**

* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41137>

List of GSM IDs :

GSM1009062 Tumor 1\_Snap Frozen

GSM1009063 Tumor 1\_4C @ 5mins

GSM1009064 Tumor 1\_4C @ 30mins

GSM1009065 Tumor 1\_4C @ 60mins

GSM1009066 Tumor 1\_4C @ 120 mins

GSM1009067 Tumor 1\_4C @ 240 mins

GSM1009068 Tumor 1\_22C @ 5 mins

GSM1009069 Tumor 1\_22C @ 30 mins

GSM1009070 Tumor 1\_22C @ 60 mins

GSM1009071 Tumor 1\_22C @ 120 mins

GSM1009072 Tumor 1\_37C @ 5 mins

GSM1009073 Tumor 1\_37C @ 30 mins

GSM1009074 Tumor 1\_37C @ 60 mins

GSM1009075 Tumor 1\_37C @ 120 mins

GSM1009076 Tumor 2\_Snap Frozen

GSM1009077 Tumor 2\_4C @ 5mins

GSM1009078 Tumor 2\_4C @ 30mins

GSM1009079 Tumor 2\_4C @ 60mins

GSM1009080 Tumor 2\_4C @ 120 mins

GSM1009081 Tumor 2\_4C @ 240 mins

GSM1009082 Tumor 2\_22C @ 5 mins

GSM1009083 Tumor 2\_22C @ 30 mins

GSM1009084 Tumor 2\_22C @ 60 mins

GSM1009085 Tumor 2\_22C @ 120 mins

GSM1009086 Tumor 2\_37C @ 5 mins

GSM1009087 Tumor 2\_37C @ 30 mins

GSM1009088 Tumor 2\_37C @ 60 mins

GSM1009089 Tumor 2\_37C @ 120 mins

GSM1009090 Tumor 3\_Snap Frozen

GSM1009091 Tumor 3\_4C @ 5mins

GSM1009092 Tumor 3\_4C @ 30mins

GSM1009093 Tumor 3\_4C @ 60mins

GSM1009094 Tumor 3\_4C @ 120 mins

GSM1009095 Tumor 3\_4C @ 240 mins

GSM1009096 Tumor 3\_22C @ 5 mins

GSM1009097 Tumor 3\_22C @ 30 mins

GSM1009098 Tumor 3\_22C @ 60 mins

GSM1009099 Tumor 3\_22C @ 120 mins

GSM1009100 Tumor 3\_37C @ 5 mins

GSM1009101 Tumor 3\_37C @ 30 mins

GSM1009102 Tumor 3\_37C @ 60 mins

GSM1009103 Tumor 3\_37C @ 120 mins

GSM1009104 Tumor 4\_Snap Frozen

GSM1009105 Tumor 4\_4C @ 5mins

GSM1009106 Tumor 4\_4C @ 30mins

GSM1009107 Tumor 4\_4C @ 60mins

GSM1009108 Tumor 4\_4C @ 120 mins

GSM1009109 Tumor 4\_4C @ 240 mins

GSM1009110 Tumor 4\_22C @ 5 mins

GSM1009111 Tumor 4\_22C @ 30 mins

GSM1009112 Tumor 4\_22C @ 60 mins

GSM1009113 Tumor 4\_22C @ 120 mins

GSM1009114 Tumor 4\_37C @ 5 mins

GSM1009115 Tumor 4\_37C @ 30 mins

GSM1009116 Tumor 4\_37C @ 60 mins

GSM1009117 Tumor 4\_37C @ 120 mins

GSM1009118 Tumor 5\_Snap Frozen

GSM1009119 Tumor 5\_4C @ 5mins

GSM1009120 Tumor 5\_4C @ 30mins

GSM1009121 Tumor 5\_4C @ 60mins

GSM1009122 Tumor 5\_4C @ 120 mins

GSM1009123 Tumor 5\_4C @ 240 mins

GSM1009124 Tumor 5\_22C @ 5 mins

GSM1009125 Tumor 5\_22C @ 30 mins

GSM1009126 Tumor 5\_22C @ 60 mins

GSM1009127 Tumor 5\_22C @ 120 mins

GSM1009128 Tumor 5\_37C @ 5 mins

GSM1009129 Tumor 5\_37C @ 30 mins

GSM1009130 Tumor 5\_37C @ 60 mins

GSM1009131 Tumor 5\_37C @ 120 mins

GSM1009132 Tumor 6\_Snap Frozen

GSM1009133 Tumor 6\_4C @ 5mins

GSM1009134 Tumor 6\_4C @ 30mins

GSM1009135 Tumor 6\_4C @ 60mins

GSM1009136 Tumor 6\_4C @ 120 mins

GSM1009137 Tumor 6\_4C @ 240 mins

GSM1009138 Tumor 6\_22C @ 5 mins

GSM1009139 Tumor 6\_22C @ 30 mins

GSM1009140 Tumor 6\_22C @ 60 mins

GSM1009141 Tumor 6\_22C @ 120 mins

GSM1009142 Tumor 6\_37C @ 5 mins

GSM1009143 Tumor 6\_37C @ 30 mins

GSM1009144 Tumor 6\_37C @ 60 mins

GSM1009145 Tumor 6\_37C @ 120 mins

GSM1009146 Tumor 7\_Snap Frozen

GSM1009147 Tumor 7\_4C @ 5mins

GSM1009148 Tumor 7\_4C @ 30mins

GSM1009149 Tumor 7\_4C @ 60mins

GSM1009150 Tumor 7\_4C @ 120 mins

GSM1009151 Tumor 7\_4C @ 240 mins

GSM1009152 Tumor 7\_22C @ 5 mins

GSM1009153 Tumor 7\_22C @ 30 mins

GSM1009154 Tumor 7\_22C @ 60 mins

GSM1009155 Tumor 7\_22C @ 120 mins

GSM1009156 Tumor 7\_37C @ 5 mins

GSM1009157 Tumor 7\_37C @ 30 mins

GSM1009158 Tumor 7\_37C @ 60 mins

GSM1009159 Tumor 7\_37C @ 120 mins

GSM1009160 Tumor 8\_Snap Frozen

GSM1009161 Tumor 8\_4C @ 5mins

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*END OF DOCUMENT\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*8