如何编写程序？

三思而后行

伊现富 (Yi Xianfu)

2017-11-18

资源共享

图 1: https://github.com/Yixf-Education/demo4bx

任务需求

- 已知：成熟 miRNA 及其 pri-miRNA 的相关信息

chr1	1169004	1169087	MI0001641_hsa-mir-429	0	+	chr1	1169054	1169076	MIMAT0001536_hsa-miR-429	1	+
chr1	1296109	1296170	MI0022571_hsa-mir-6726	0	-	chr1	1296144	1296165	MIMAT0027353_hsa-miR-6726-5p	1	-
chr1	1296109	1296170	MI0022571_hsa-mir-6726	0	-	chr1	1296109	1296129	MIMAT0027354_hsa-miR-6726-3p	1	-
chr1	1312501	1312566	MI0022572_hsa-mir-6727	0	-	chr1	1312538	1312561	MIMAT0027355_hsa-miR-6727-5p	1	-
chr1	1312501	1312566	MI0022572_hsa-mir-6727	0	-	chr1	1312501	1312521	MIMAT0027356_hsa-miR-6727-3p	1	-
chr1	1339649	1339708	MI0022653_hsa-mir-6808	0	-	chr1	1339681	1339703	MIMAT0027516_hsa-miR-6808-5p	1	-
chr1	1339649	1339708	MI0022653_hsa-mir-6808	0	-	chr1	1339649	1339670	MIMAT0027517_hsa-miR-6808-3p	1	-
chr1	3127974	3128035	MI0015861_hsa-mir-4251	0	+	chr1	3128008	3128025	MIMAT0016883_hsa-miR-4251	1	+
chr1	3560694	3560790	MI0003556_hsa-mir-551a	0	-	chr1	3560709	3560730	MIMAT0003214_hsa-miR-551a	1	-
chr1	5564070	5564143	MI0016753_hsa-mir-4417	0	+	chr1	5564081	5564099	MIMAT0018929_hsa-miR-4417	1	+
chr1	5862671	5862741	MI0017322_hsa-mir-4689	0	-	chr1	5862710	5862732	MIMAT0019778_hsa-miR-4689	1	-
chr1	6429833	6429896	MI0015864_hsa-mir-4252	0	-	chr1	6429843	6429862	MIMAT0016886_hsa-miR-4252	1	-
chr1	8866501	8866590	MI0022573_hsa-mir-6728	0	-	chr1	8866562	8866585	MIMAT0027357_hsa-miR-6728-5p	1	-
chr1	8866501	8866590	MI0022573_hsa-mir-6728	0	-	chr1	8866501	8866522	MIMAT0027358_hsa-miR-6728-3p	1	-
chr1	9151667	9151777	MI0000268_hsa-mir-34a	0	-	chr1	9151734	9151756	MIMAT0000255_hsa-miR-34a-5p	1	-
chr1	9151667	9151777	MI0000268_hsa-mir-34a	0	-	chr1	9151692	9151714	MIMAT0004557_hsa-miR-34a-3p	1	-
chr1	9967380	9967458	MI0019304_hsa-mir-5697	0	+	chr1	9967389	9967411	MIMAT0022490_hsa-miR-5697	1	+
chr1	10227717	10227803	MI0014254_hsa-mir-1273d	0	+	chr1	1E+07	10227751	MIMAT0015090_hsa-miR-1273d	1	+

图 2:

- 要求：按照位置的重要性将 pri-miRNA 分段
- 实现：编写程序 …… …… 如何编写程序？

了解背景知识：miRNA

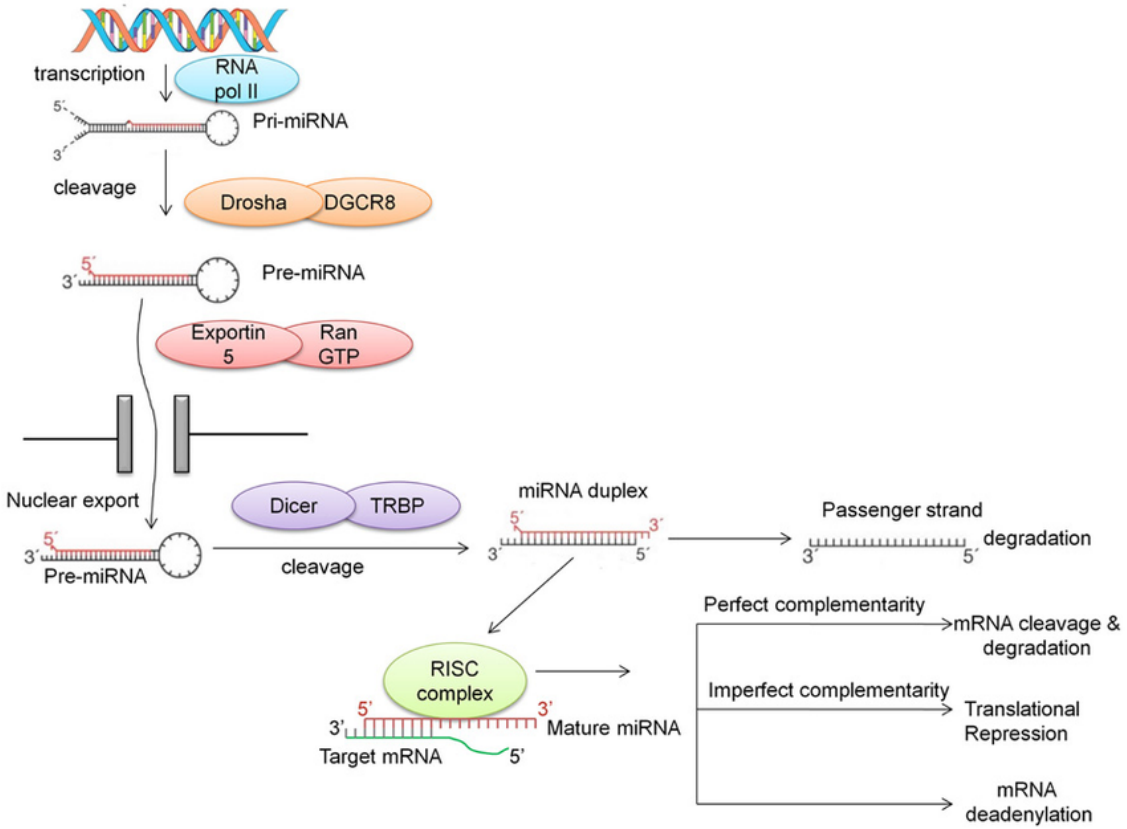


图 3:

了解背景知识：miRNA 的产生过程——pri2pre

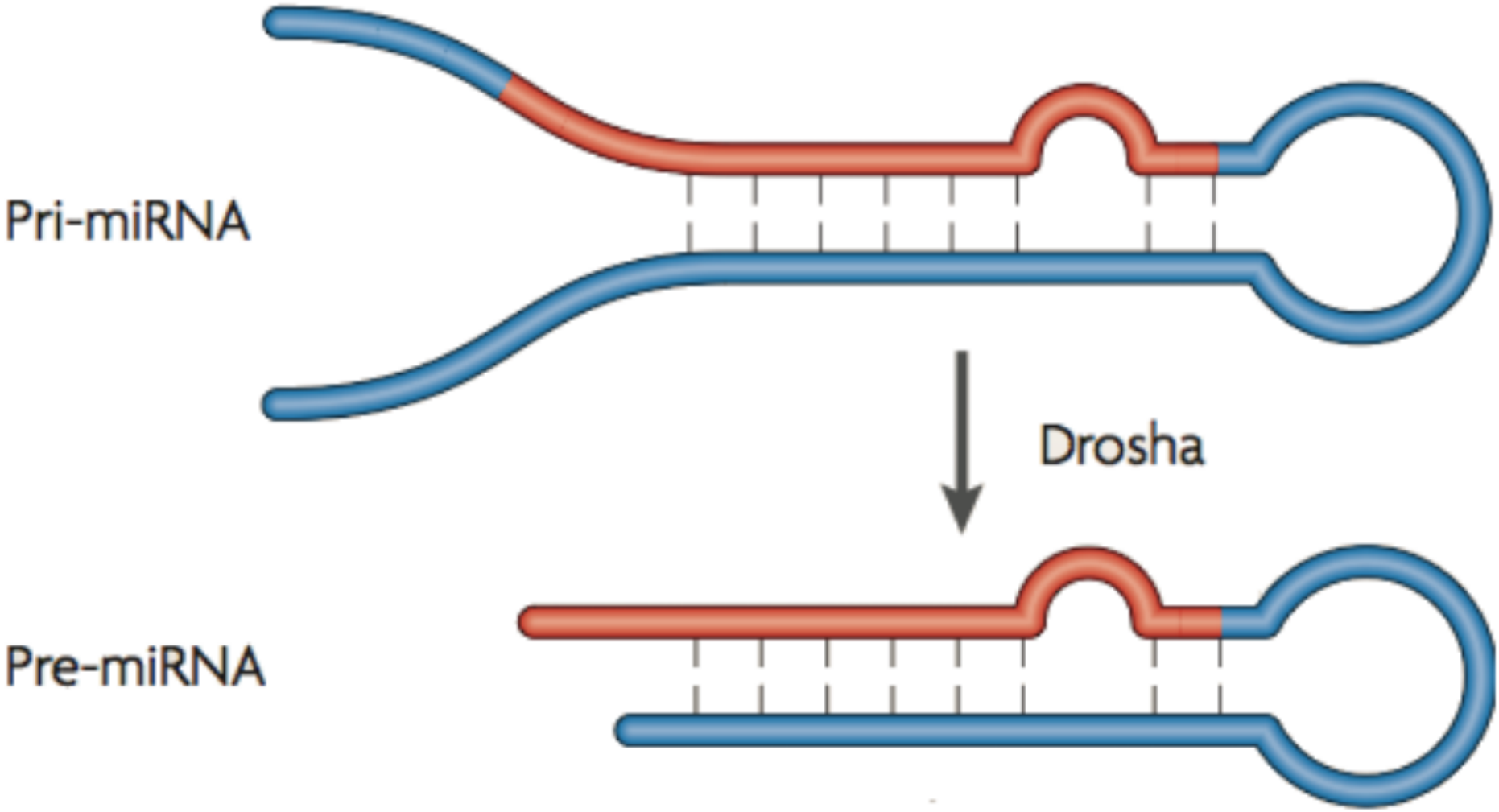


图 4:

了解背景知识：miRNA 的产生过程——pre2mature

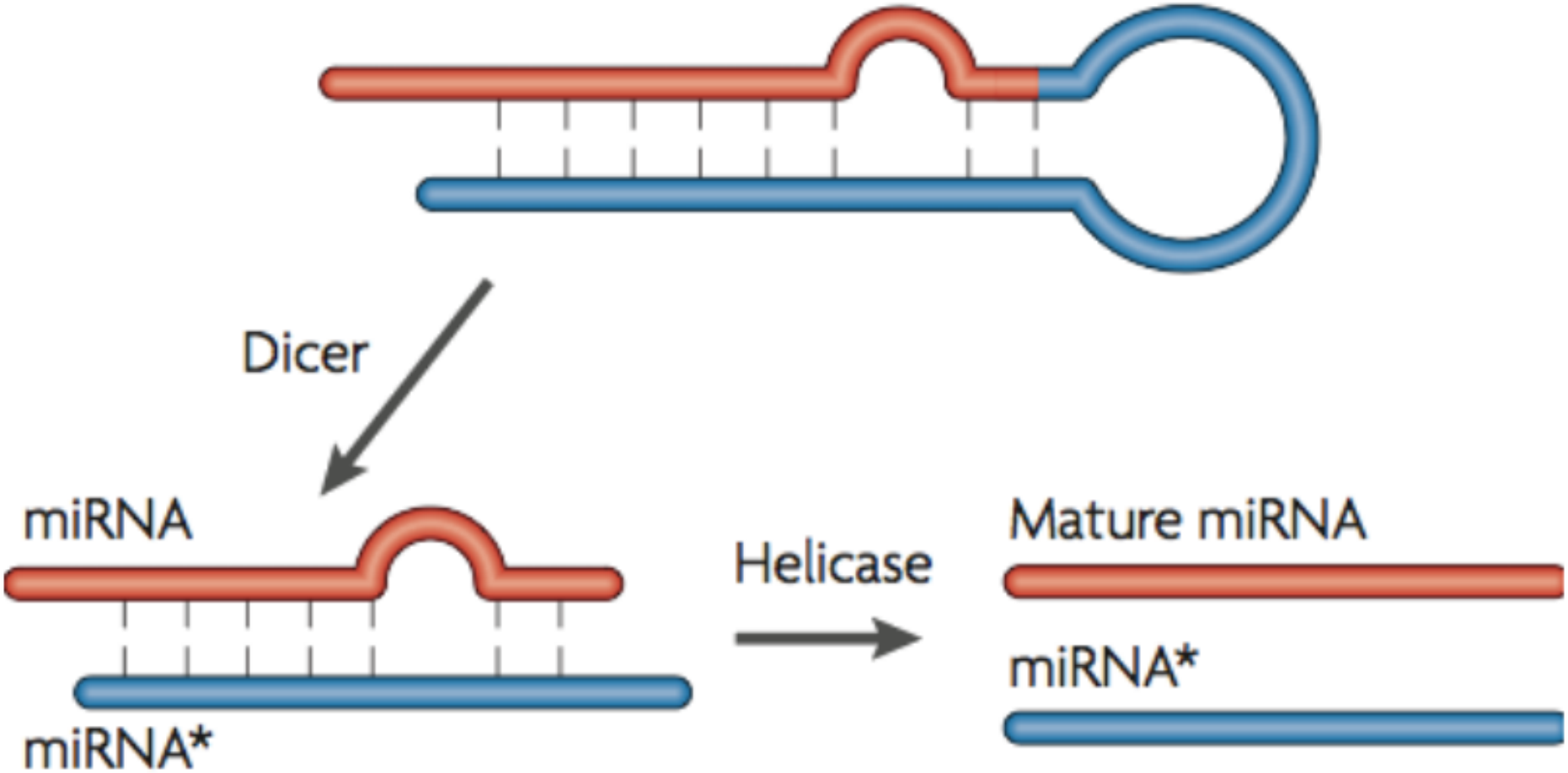


图 5:

了解背景知识：miRNA 的作用方式——种子区域

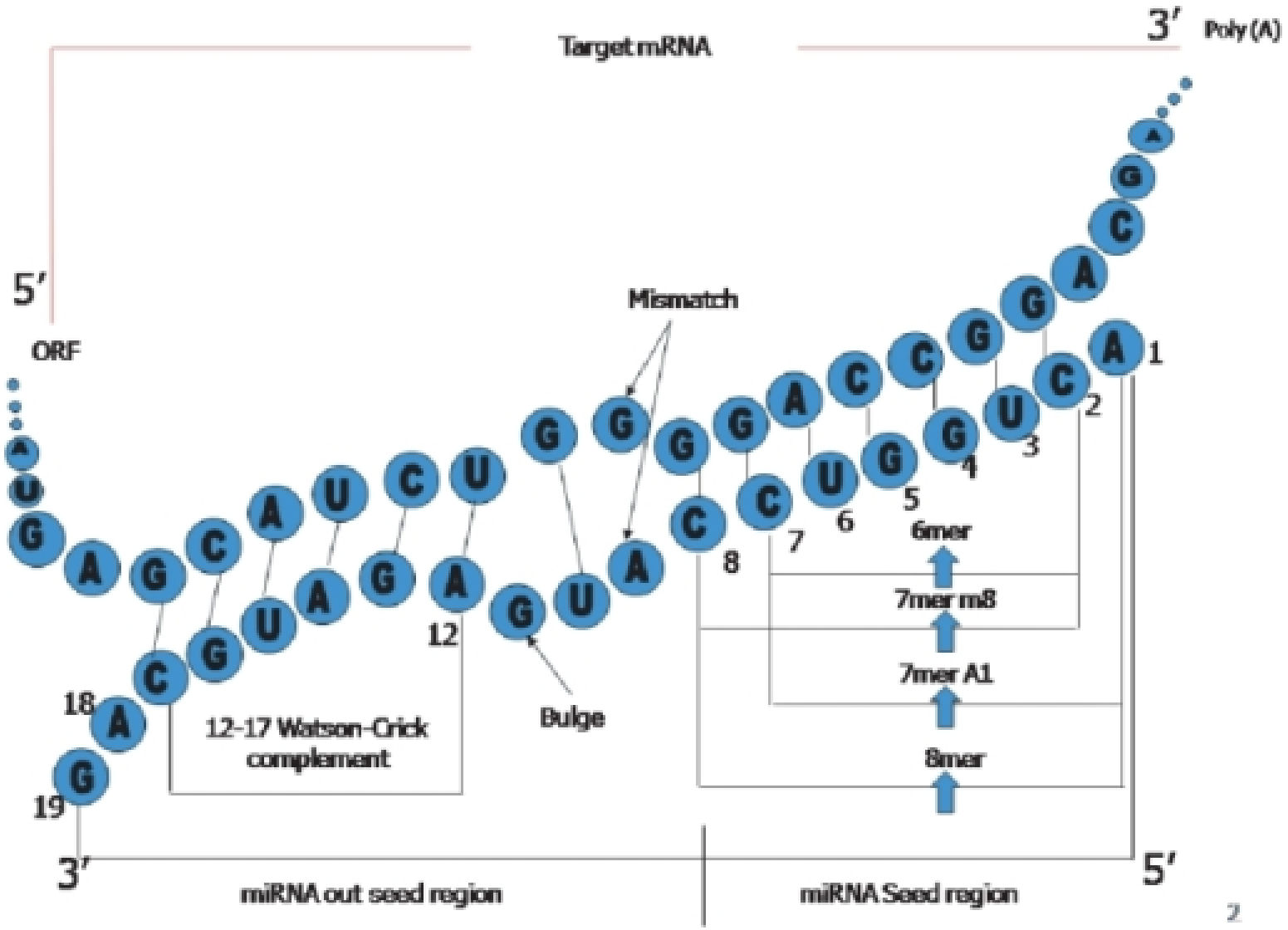


图 6:

任务需求的具体化——辅助示意图

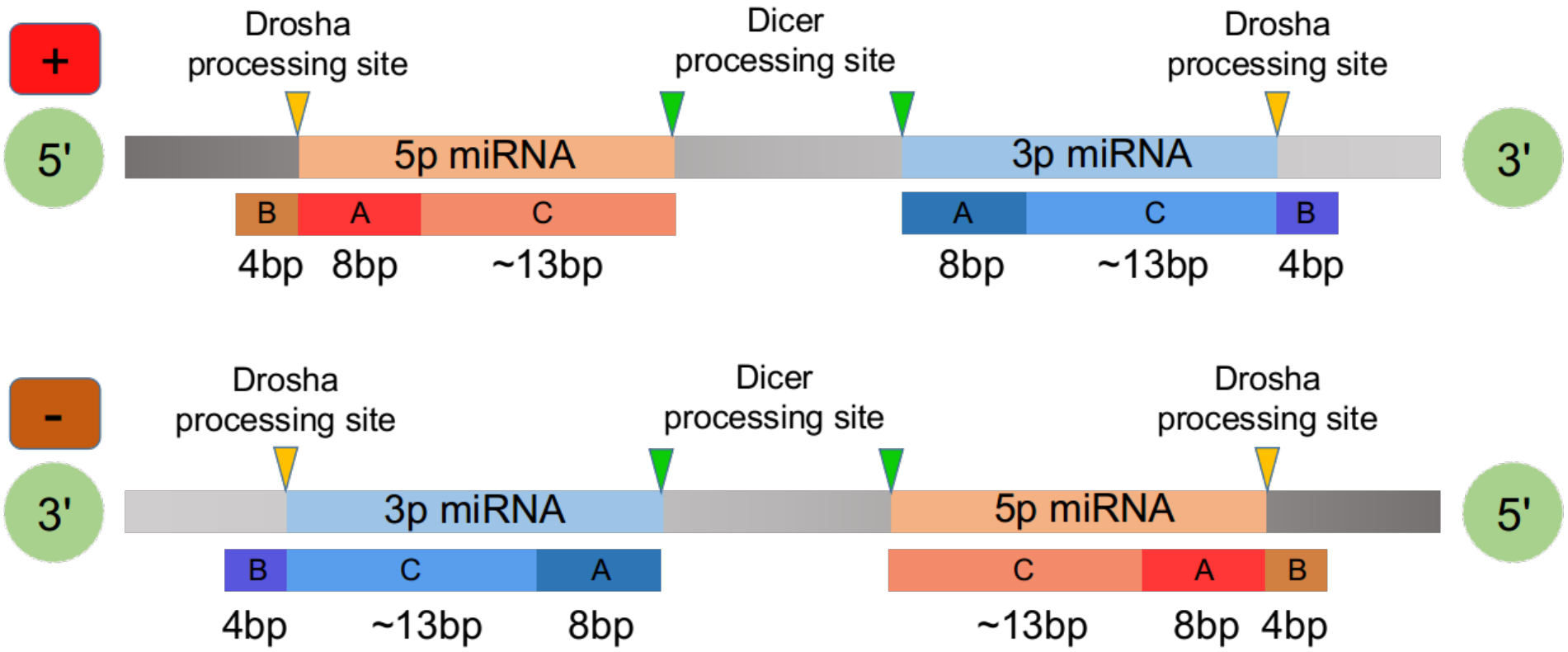


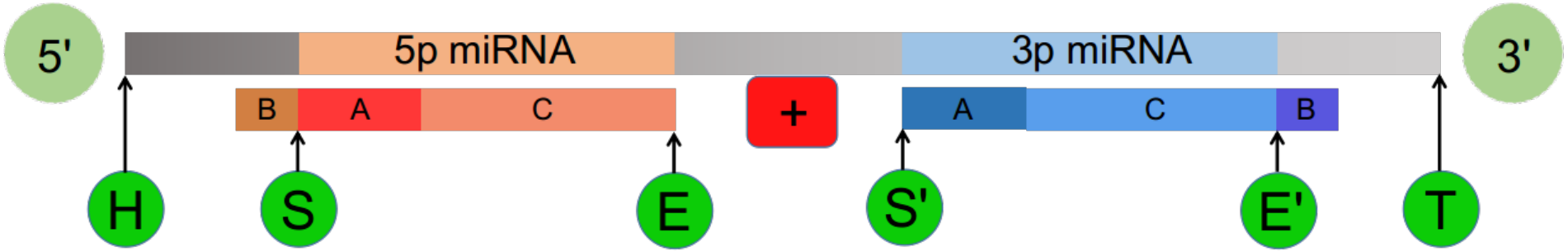
图 7:

理解输入数据：文件格式及数据信息

```
chr8    1801124 1801192 MI0016075_hsa-mir-3674  0   +   chr8    1801132 1801154 MIMAT0018097_hsa-miR-3674   1   +
chr8    1817230 1817307 MI0003608_hsa-mir-596   0   +   chr8    1817245 1817266 MIMAT0003264_hsa-miR-596    1   +
chr8    2076588 2076640 MI0023621_hsa-mir-7160  0   +   chr8    2076588 2076609 MIMAT0028230_hsa-miR-7160-5p    1   +
chr8    2076588 2076640 MI0023621_hsa-mir-7160  0   +   chr8    2076619 2076640 MIMAT0028231_hsa-miR-7160-3p    1   +
chr8    6622123 6622220 MI0025891_hsa-mir-8055  0   -   chr8    6622133 6622156 MIMAT0030982_hsa-miR-8055   1   -
chr8    6745163 6745244 MI0017287_hsa-mir-4659a 0   +   chr8    6745176 6745198 MIMAT0019726_hsa-miR-4659a-5p   1   +
chr8    6745163 6745244 MI0017287_hsa-mir-4659a 0   +   chr8    6745211 6745233 MIMAT0019727_hsa-miR-4659a-3p   1   +
chr8    6745163 6745244 MI0017287_hsa-mir-4659a 0   +   chr8    6745212 6745231 MIMAT0019733_hsa-miR-4659b-5p   1   -
chr8    6745163 6745244 MI0017287_hsa-mir-4659a 0   +   chr8    6745174 6745196 MIMAT0019734_hsa-miR-4659b-3p   1   -
chr8    6745167 6745240 MI0017291_hsa-mir-4659b 0   -   chr8    6745176 6745198 MIMAT0019726_hsa-miR-4659a-5p   1   +
chr8    6745167 6745240 MI0017291_hsa-mir-4659b 0   -   chr8    6745211 6745233 MIMAT0019727_hsa-miR-4659a-3p   1   +
chr8    6745167 6745240 MI0017291_hsa-mir-4659b 0   -   chr8    6745212 6745231 MIMAT0019733_hsa-miR-4659b-5p   1   -
chr8    6745167 6745240 MI0017291_hsa-mir-4659b 0   -   chr8    6745174 6745196 MIMAT0019734_hsa-miR-4659b-3p   1   -
chr8    8088940 8089089 MI0006423_hsa-mir-548i-3    0   -   chr8    8089029 8089051 MIMAT0005935_2_hsa-miR-548i 1   -
chr8    9048444 9048518 MI0017288_hsa-mir-4660  0   +   chr8    9048452 9048475 MIMAT0019728_hsa-miR-4660   1   +
chr8    9741671 9741768 MI0003609_hsa-mir-597   0   +   chr8    9741686 9741708 MIMAT0003265_hsa-miR-597-5p  1   +
chr8    9741671 9741768 MI0003609_hsa-mir-597   0   +   chr8    9741728 9741751 MIMAT0026619_hsa-miR-597-3p  1   +
chr8    9903387 9903472 MI0000443_hsa-mir-124-1 0   -   chr8    9903437 9903459 MIMAT0004591_1_hsa-miR-124-5p   1   -
chr8    9903387 9903472 MI0000443_hsa-mir-124-1 0   -   chr8    9903400 9903420 MIMAT0000422_1_hsa-miR-124-3p   1   -
```

图 8:

整合输入数据与任务需求——实例解析

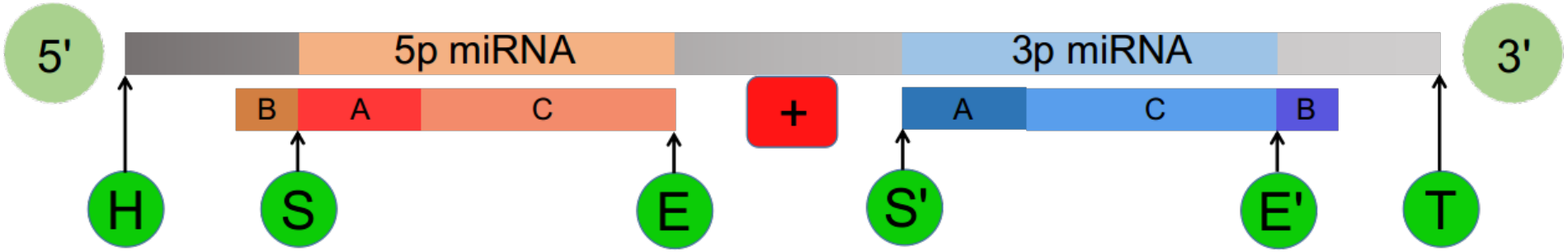


hsa-mir-200b (+)			
5p miRNA		3p miRNA	
H	1167103	T	1167198
S	1167123	S'	1167159
E	1167145	E'	1167181

Regions	5p miRNA		3p miRNA	
A	1167123	1167130	1167159	1167166
B	1167119	1167122	1167182	1167185
C	1167131	1167145	1167167	1167181

图 9:

整合输入数据与任务需求——从实例到规律



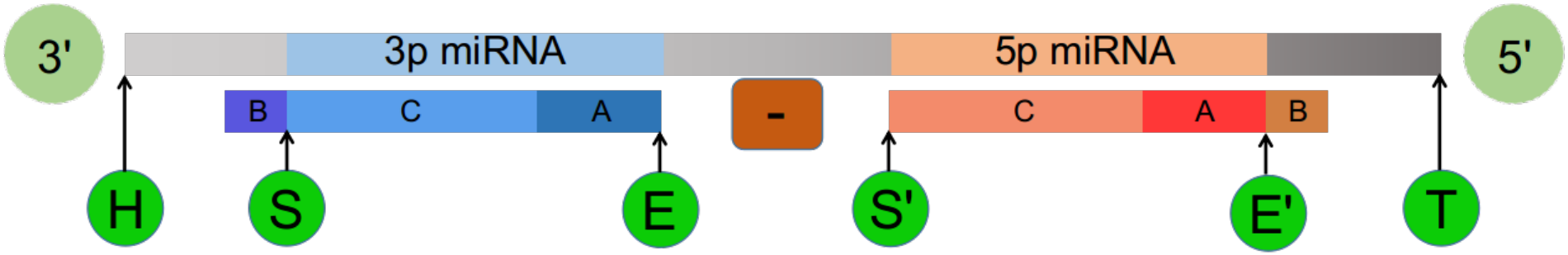
hsa-mir-200b (+)			
5p miRNA		3p miRNA	
H	1167103	T	1167198
S	1167123	S'	1167159
E	1167145	E'	1167181

Regions	5p miRNA		3p miRNA	
A	1167123	1167130	1167159	1167166
B	1167119	1167122	1167182	1167185
C	1167131	1167145	1167167	1167181

5p miRNA		3p miRNA	
S	S+7	S'	S'+7
S-4	S-1	E'+1	E'+4
S+8	E	S'+8	E'

图 10:

整合输入数据与任务需求——实例解析（续）

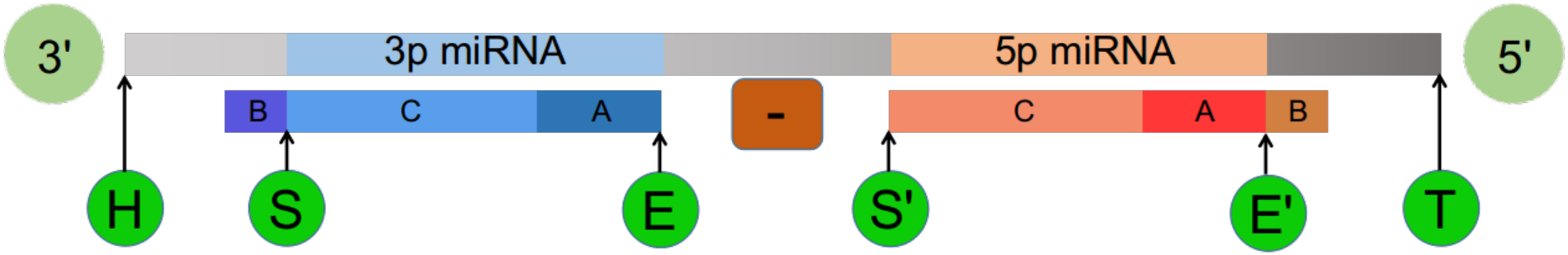


hsa-mir-6859-1 (-)			
3p miRNA		5p miRNA	
H	17368	T	17436
S	17368	S'	17408
E	17391	E'	17431

Regions	3p miRNA		5p miRNA	
A	17384	17391	17424	17431
B	17364	17367	17432	17435
C	17368	17383	17408	17423

图 11:

整合输入数据与任务需求——从实例到规律（续）



hsa-mir-6859-1 (-)			
3p miRNA		5p miRNA	
H	17368	T	17436
S	17368	S'	17408
E	17391	E'	17431

Regions	3p miRNA		5p miRNA	
A	17384	17391	17424	17431
B	17364	17367	17432	17435
C	17368	17383	17408	17423

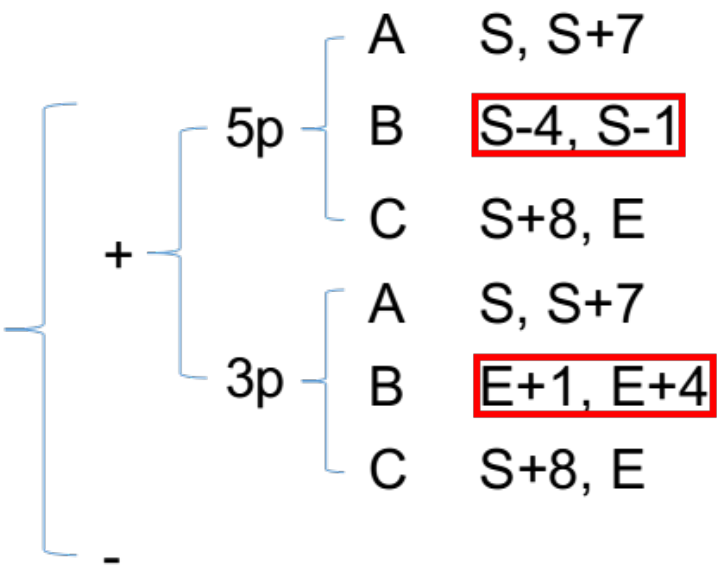
3p miRNA		5p miRNA	
E-7	E	E'-7	E'
S-4	S-1	E'+1	E'+4
S	E-8	S'	E'-8

图 12:

分析规律并进行整合——从规律到逻辑

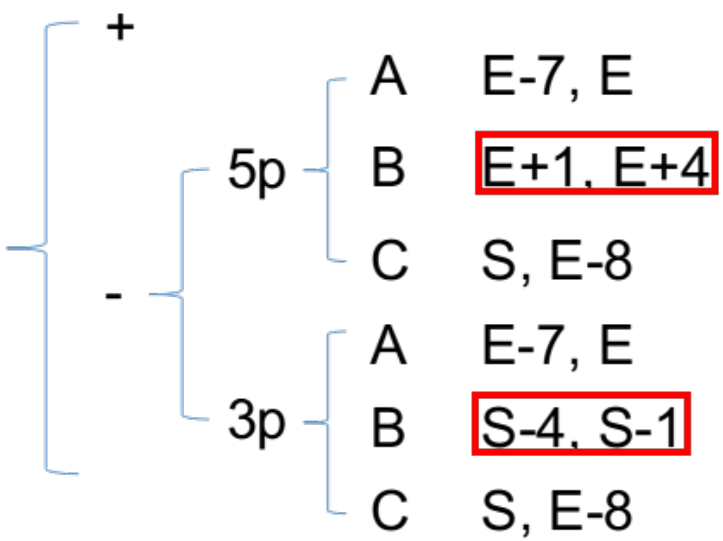
+

5p miRNA		3p miRNA	
S	S+7	S'	S'+7
S-4	S-1	E'+1	E'+4
S+8	E	S'+8	E'



-

3p miRNA		5p miRNA	
E-7	E	E'-7	E'
S-4	S-1	E'+1	E'+4
S	E-8	S'	E'-8



根据逻辑设计程序——从逻辑到设计

1. 获取正负链信息
   - 提取特定列
2. 获取或确定 5p/3p 信息
   - 模式匹配 5p/3p
   - +/-, 成熟 miRNA 的中点 vs. pri-miRNA 的中点
3. 计算 A/B/C 区域的位置
   - +/-, 5p/3p, A/B/C

根据程序设计编写伪代码——从设计到伪代码

```
foreach line
    strand = get_strand
    ps = define_ps
    if strand is "+"
        cal_plus
    if strand is "-"
        cal_minus

sub cal_plus
    cal_plus_AandC
    cal_plus_B

sub cal_minus
    cal_minus_AandC
    cal_minus_B
```

根据伪代码设计子程序——从伪代码到子程序

```
get_strand
    in: line or certain column
    out: "+" or "-"

define_ps
    in: line or certain columns
    out: "5p" or "3p"

cal_plus/minus
    in: ps and S and E
    calculate A,C
    B = call cal_plus/minus_B
    out: coordinate of A,B,C

cal_plus/minus_B
    in: ps and S and E
    out: coordinate of B
```

编写最终的程序——从思考到编程

- 编写程序并进行调试
  1. 规划好最终的输出形式（行，列，附加信息，等）
  2. 用自己熟悉的编程语言编写程序
  3. 对照实例结果校验程序输出
- 如果还是无法编写出程序
  - 没有真正理解程序设计
    - 没有讲解清楚
    - 没有真正听懂
  - 对使用的编程语言不熟悉
- 每次编程都要进行这么繁琐的设计吗？
  - 答案是肯定的！——自己理不清，电脑更糊涂
  - 答案是否定的！——大事可化小，小事可化了