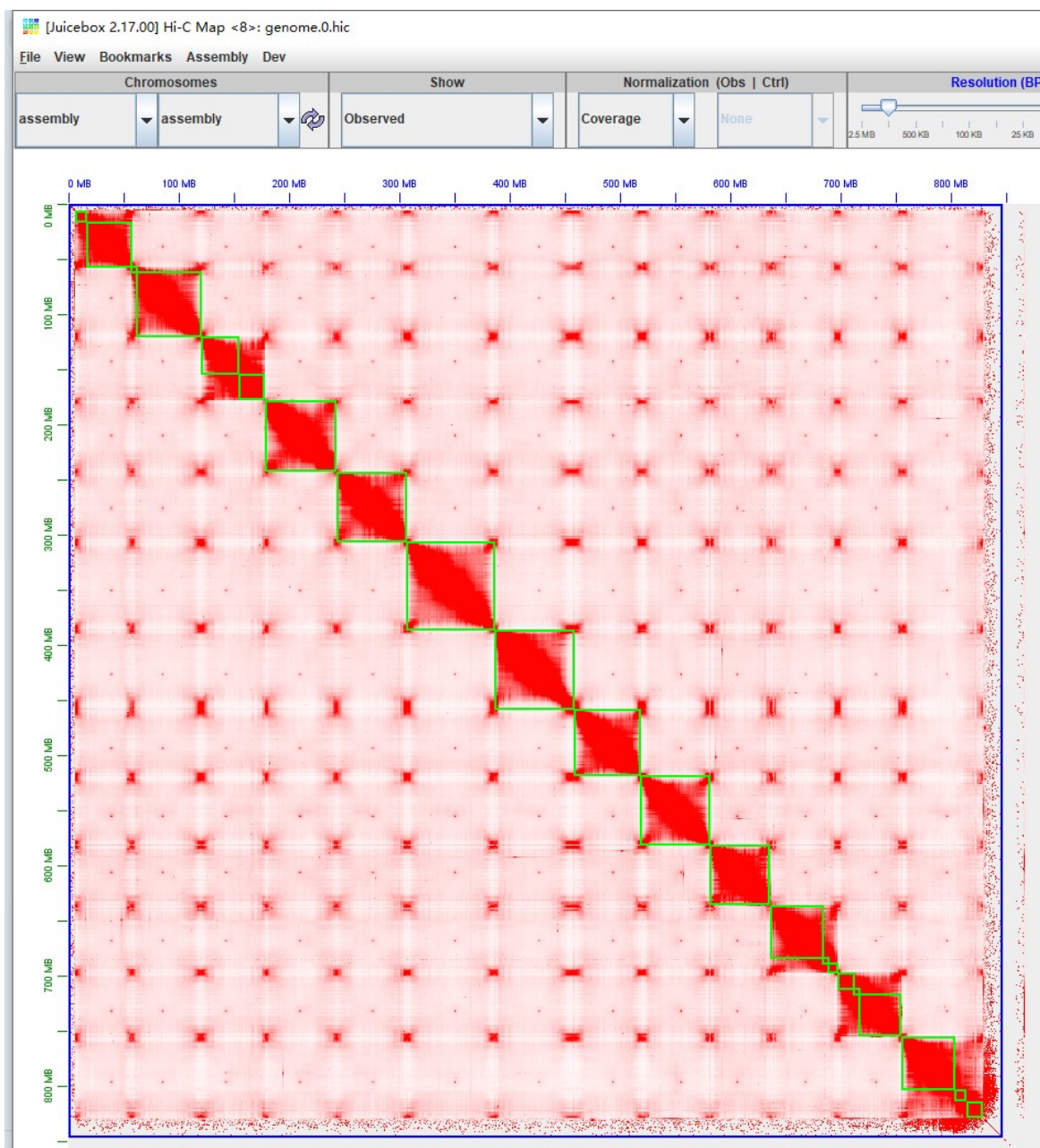
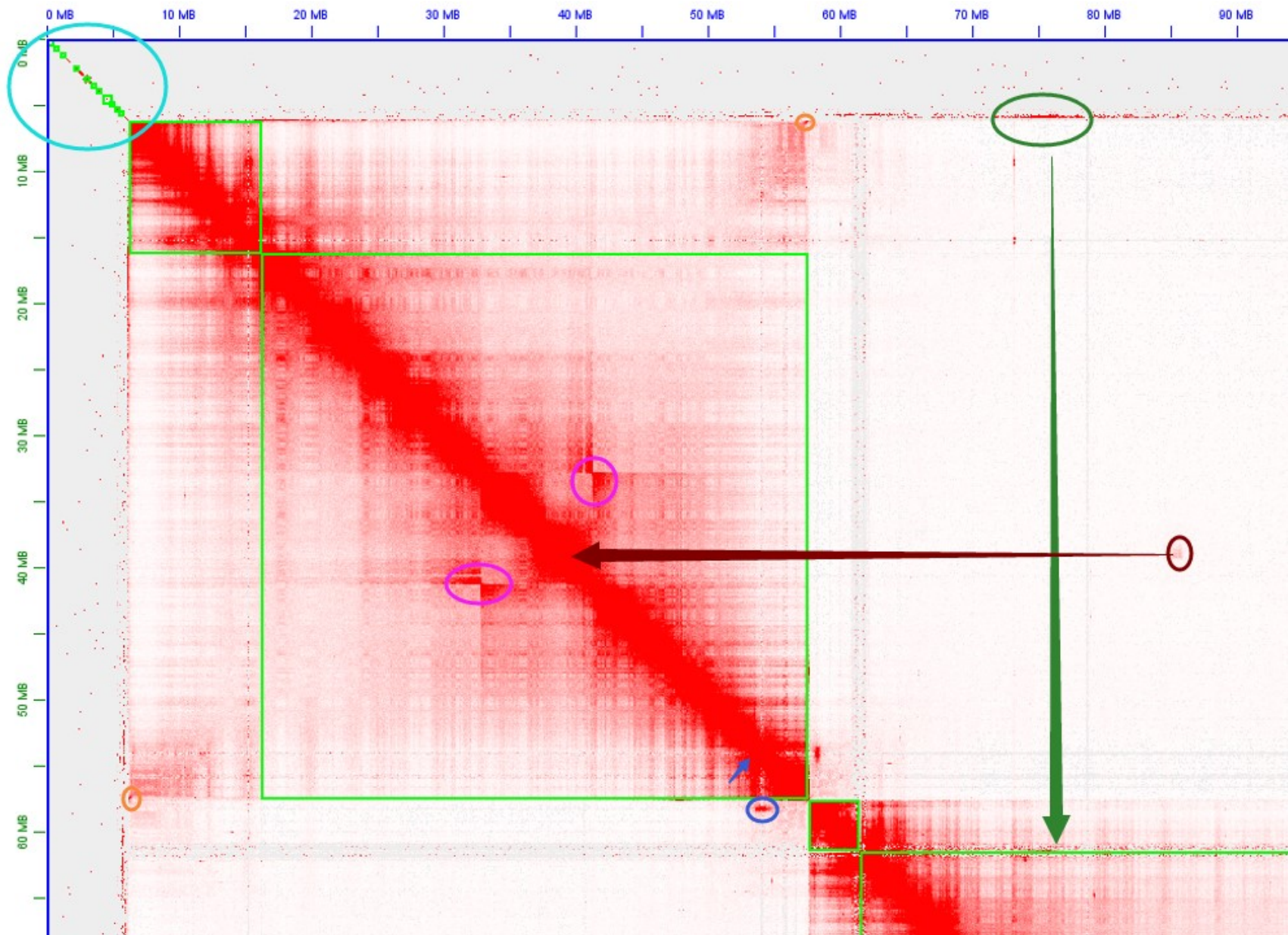


# HIC调图讲解



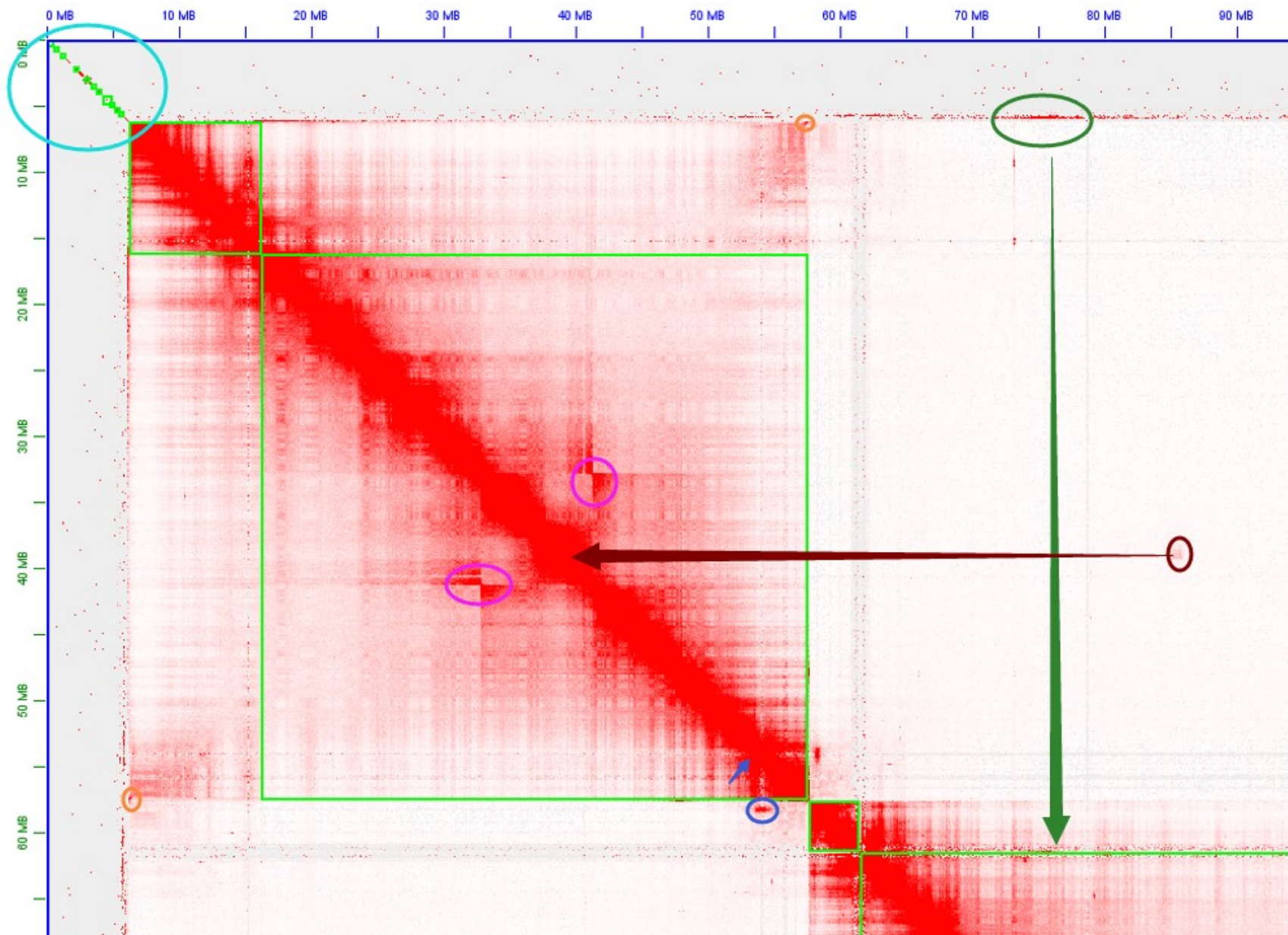
初步来看，大部分染色体都是由一个contig组成的，可见连续性还是可以的。

部分染色体（如，左上的第一条染色体和右下角的最后一条染色体）存在于一些细小的contig有红色的HIC互动信号，推测其原来可能是一段很长的重复区域，由于没有组装得很好，而细小地散落掉了。



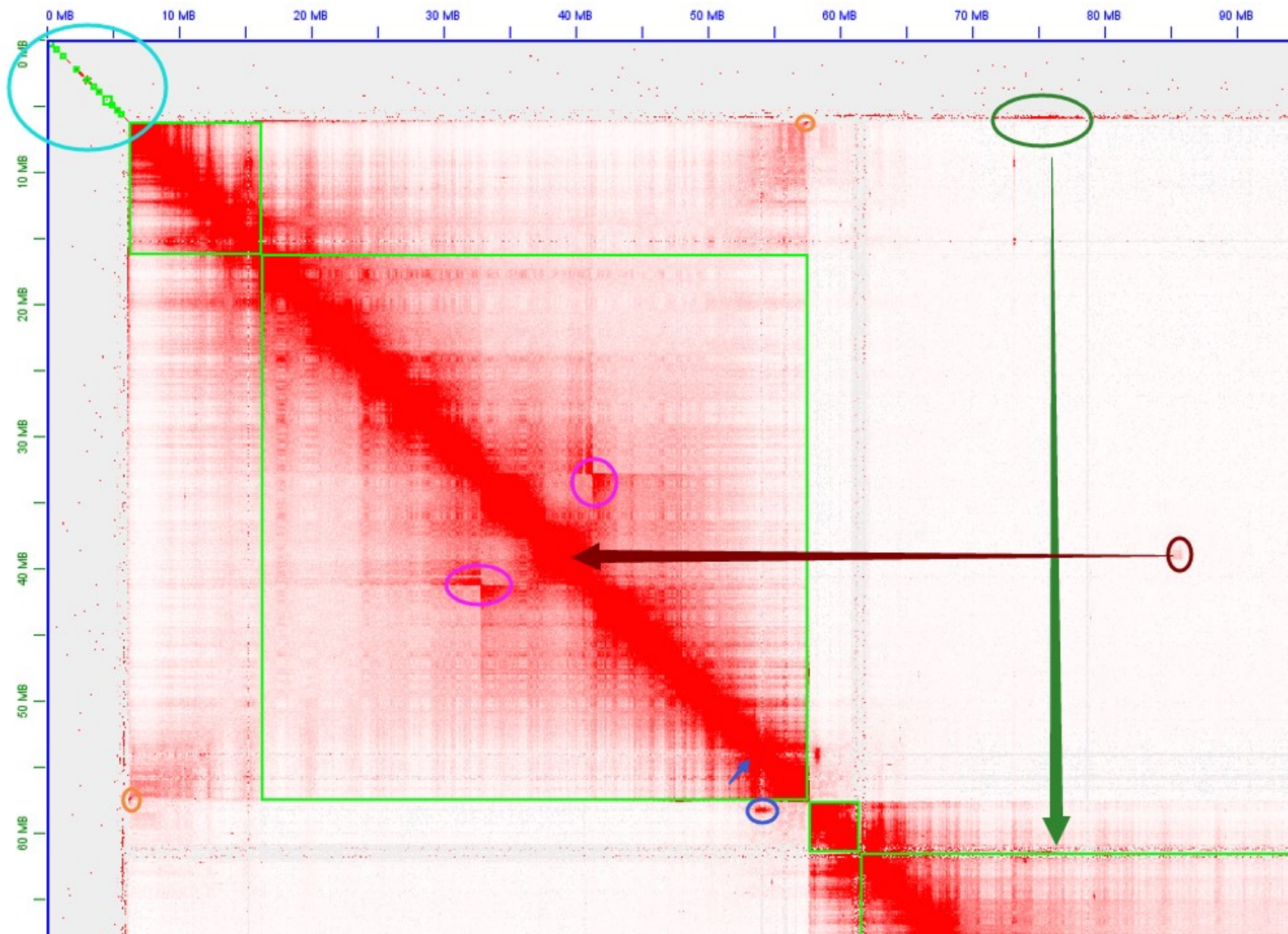
①左上角蓝圈部分是小ctg，其HIC互作信号基本都是空白的或者只有一点点红色，这是因为这些一般都是重复序列，空白的原因一方面是序列内部高度重复，二是染色体部分已经存在于该部分相似的序列。对于这些小ctg建议移到未挂账区。





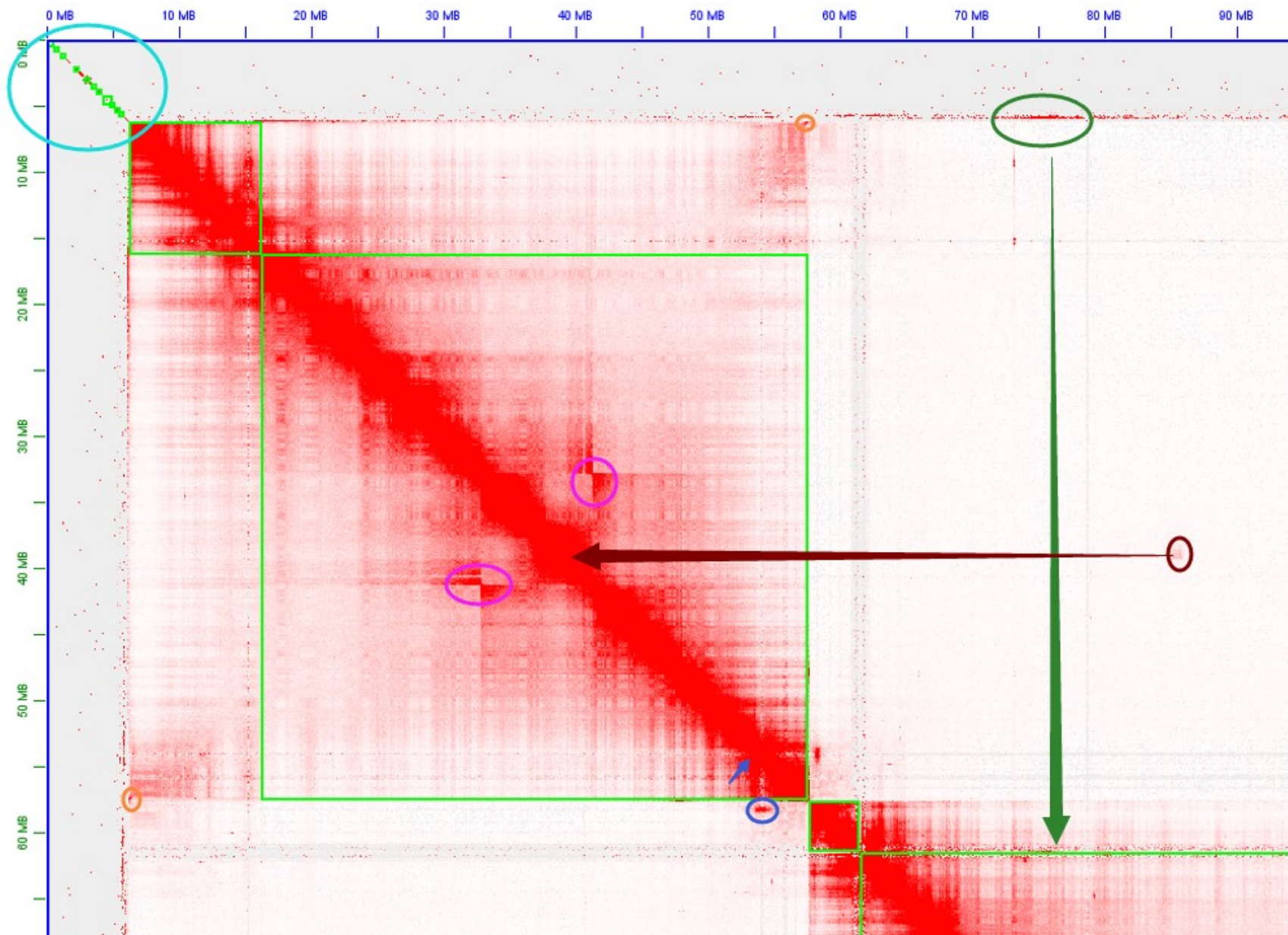
②绿圈部分存在红色的互作信号，检查发现是蓝圈部分的某个小ctg产生，对于互作信号，选择信号最强烈的部分，右键跳转对应的位置，发现是下一条染色体的某个contig的内部，而不是两个contig之间的Gap。对于完整的contig我们不会选择切开，强行把序列插入进去，故绿圈信号所示的小ctg应该被移到未挂载区。





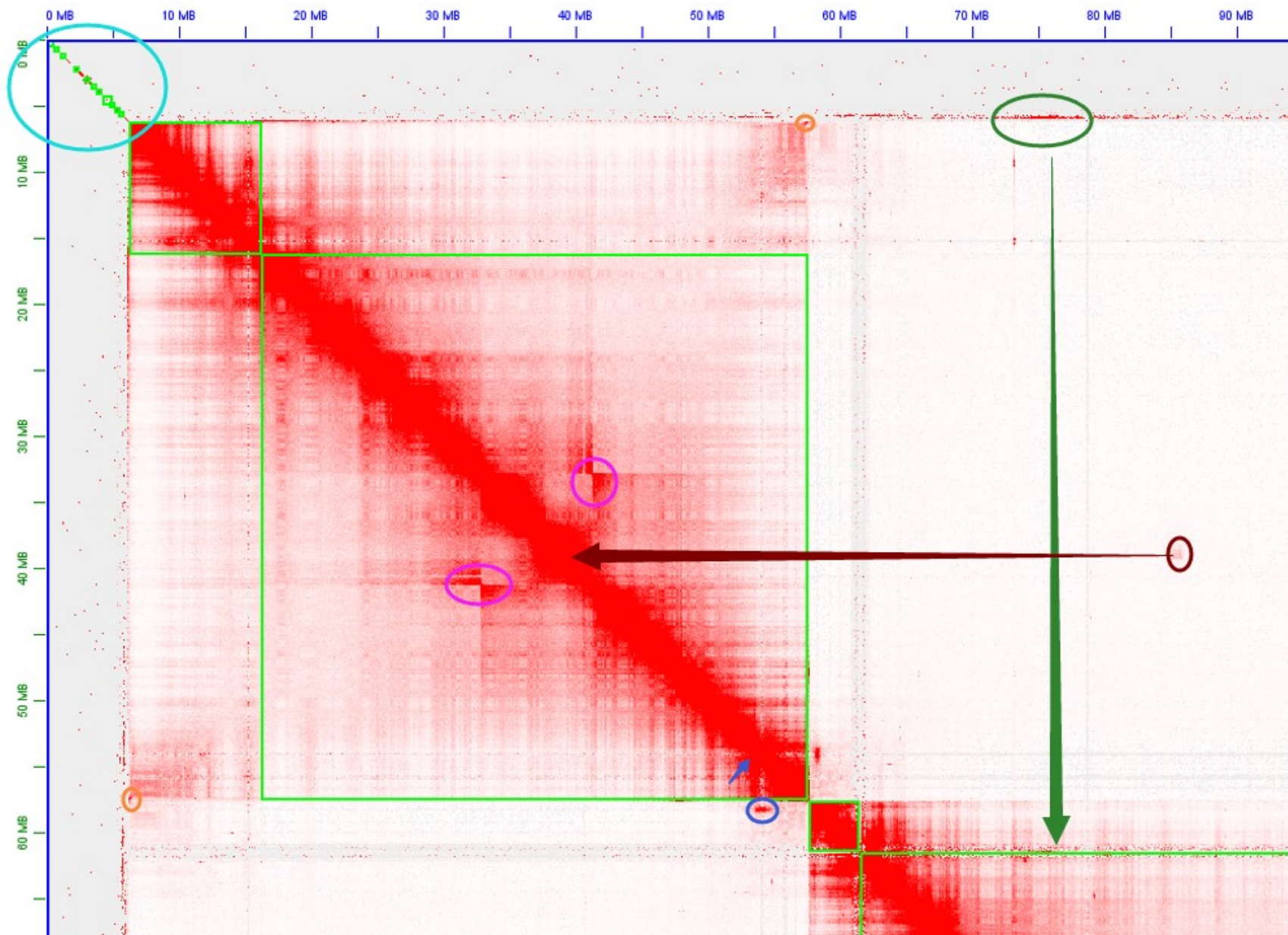
③每个染色体的左下角和右上角一般都会有一个小红点，这是因为末端存在端粒序列，6-7bp的端粒基序在高度重复，一般可达几千次重复，所以在HIC调图的时候可以通过该点来看染色体是否有端粒。棕色部分则是高度重复的一个标志，比端粒互作要大一点，实际上着丝粒的互作信号，即棕色箭头所指大致是端粒区域。





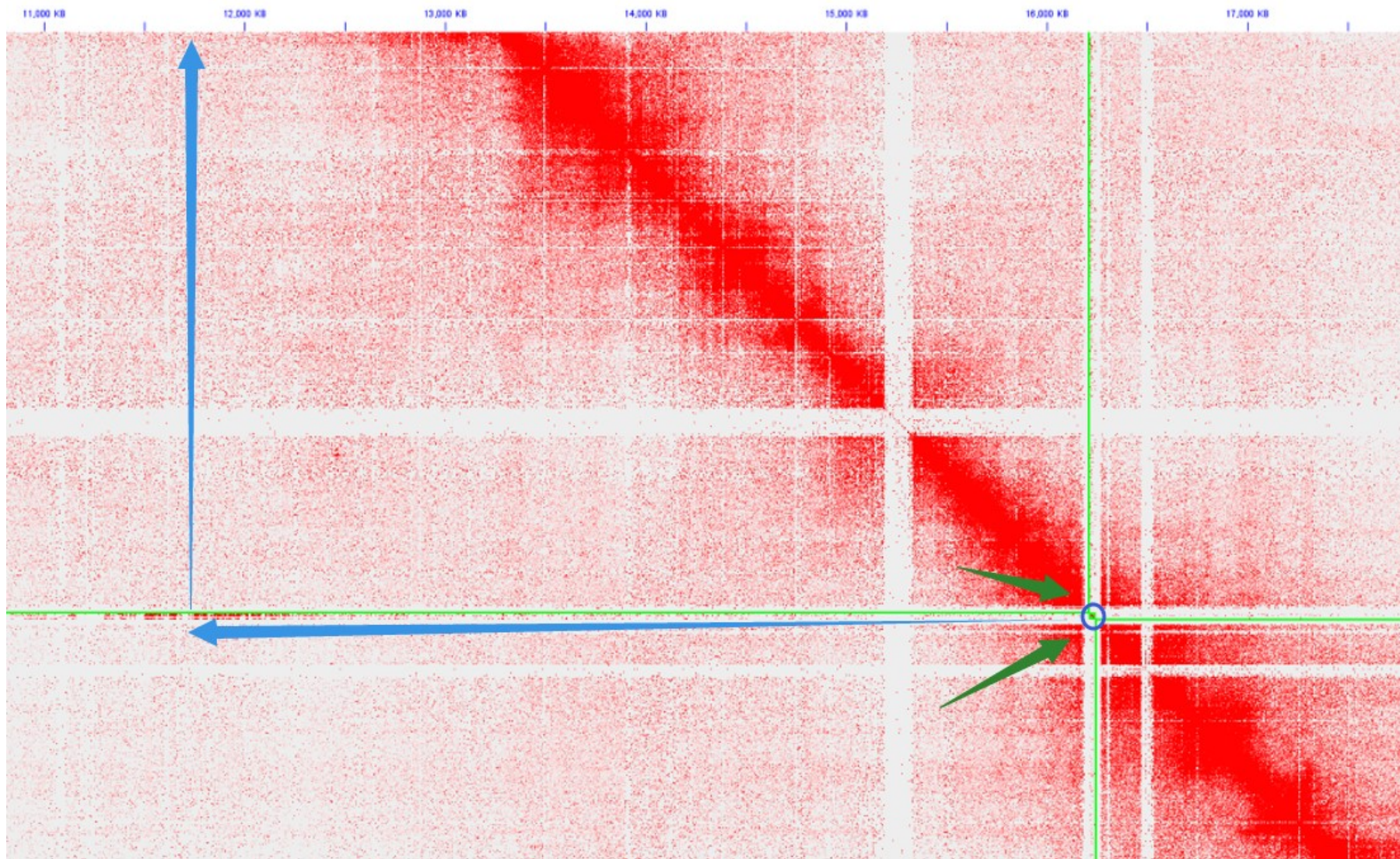
④右下角的深蓝圈，其互作信号也是重复序列信号，可能是LTR的富集区，对于以LTR/Gypsy和LTR/Copia为主的着丝粒，其往往也标识着丝粒区。





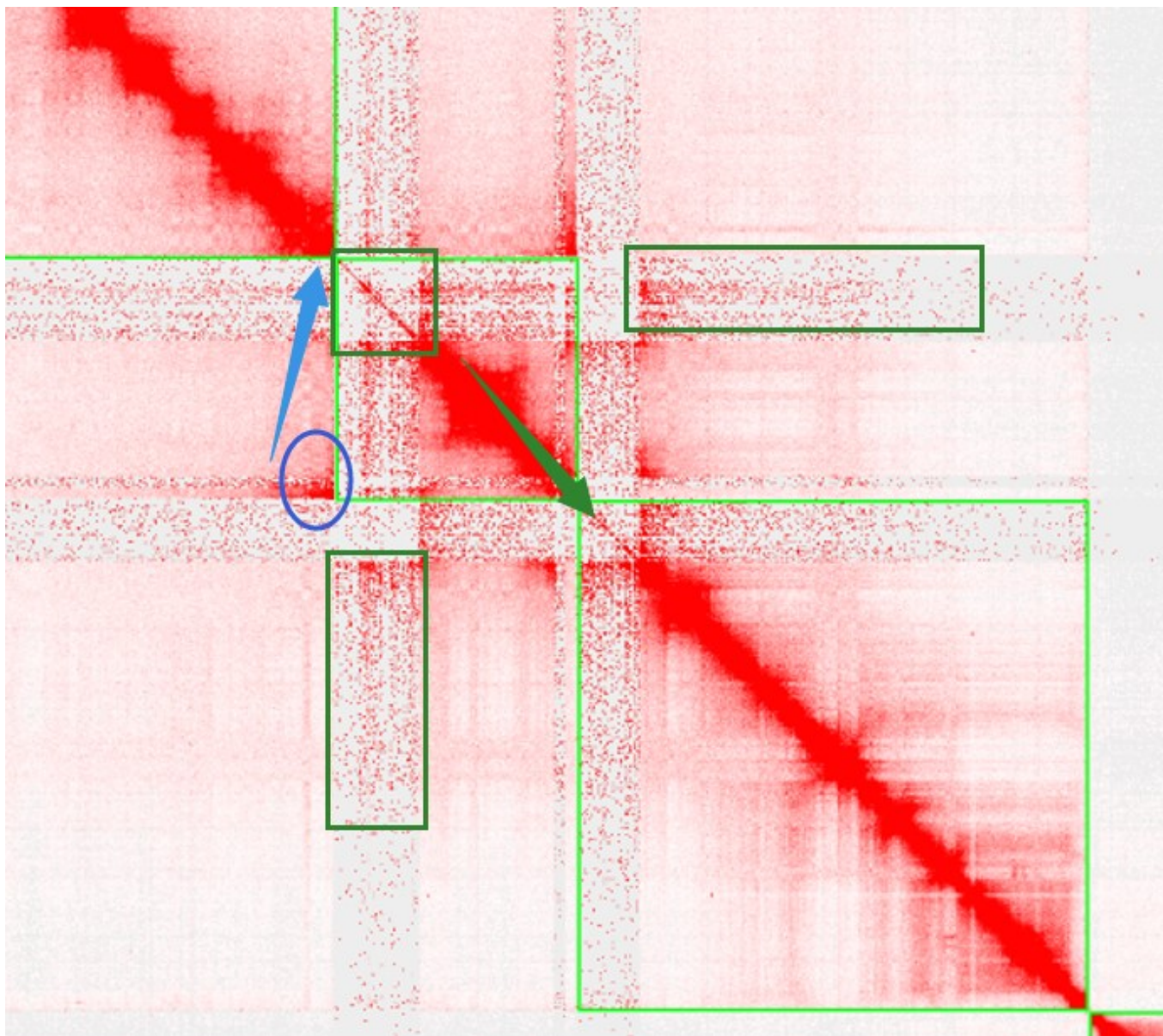
⑤紫圈部分则是HIC信号告诉的问题，一般来说我们要把红色往里调图，而该区域却没有和互动强烈的部分挨在一起，将该ctg切开后，再反转也是一样，实际上这是该区域杂合度比较高的表现。通过primary/hap1/hap2三个部分比较该区域，会发现是有个hap存在SV，实际我们不用切开该区域，保留即可。





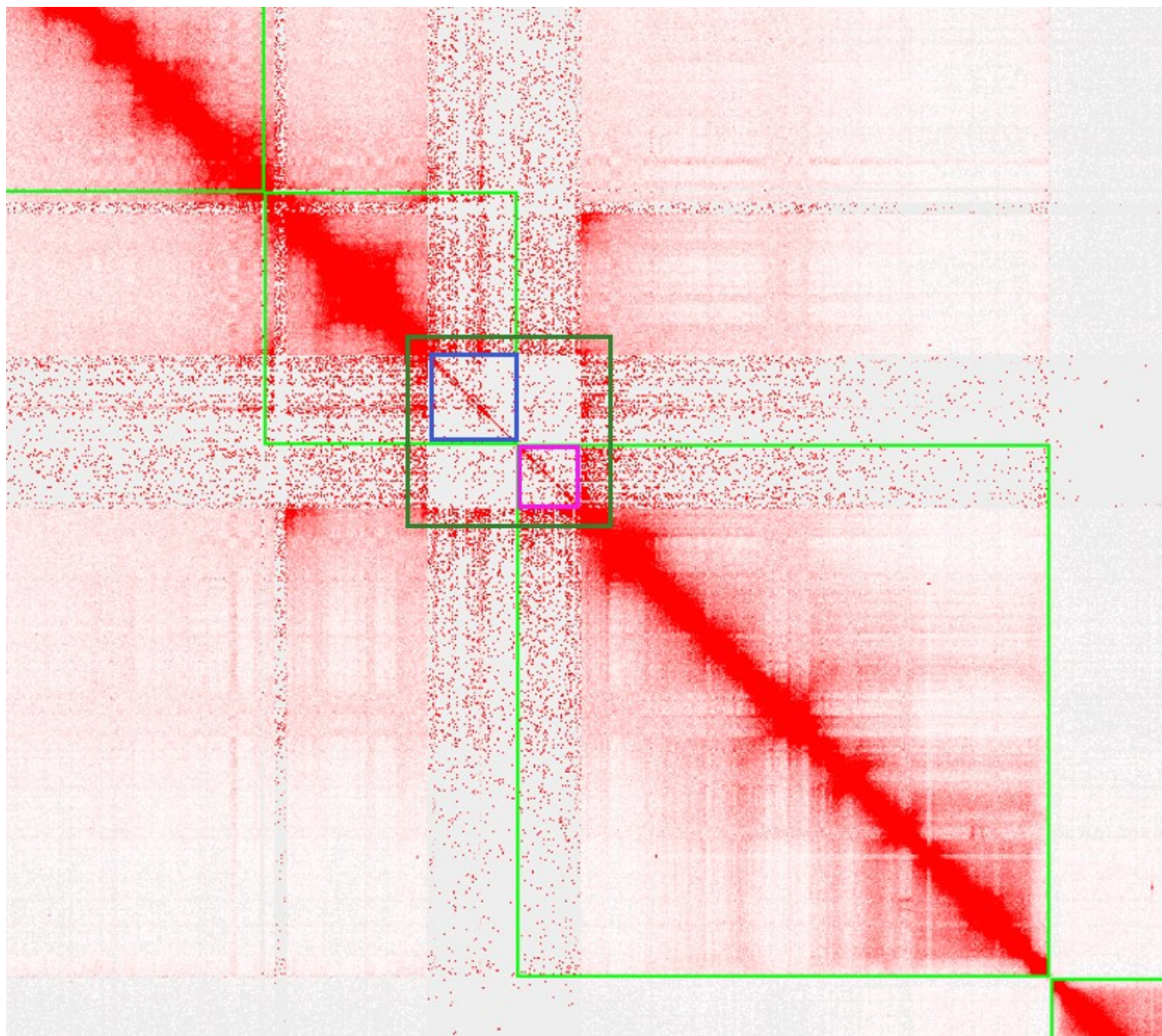
了解了以上的原理后，再来看一些常规的HIC调图部分：上图的绿色箭头所指区域，互作强烈应该挨在一起，但是中间夹了一个蓝圈的小ctg，其本身有互作信号，但是不是在当前这个位置，实际在蓝色箭头指的互作信号比较红的地方，而再往上看，箭头所指的是一个完整的长contig，因此认为该小ctg是上一个大ctg的杂合序列，因此移到未挂载区。





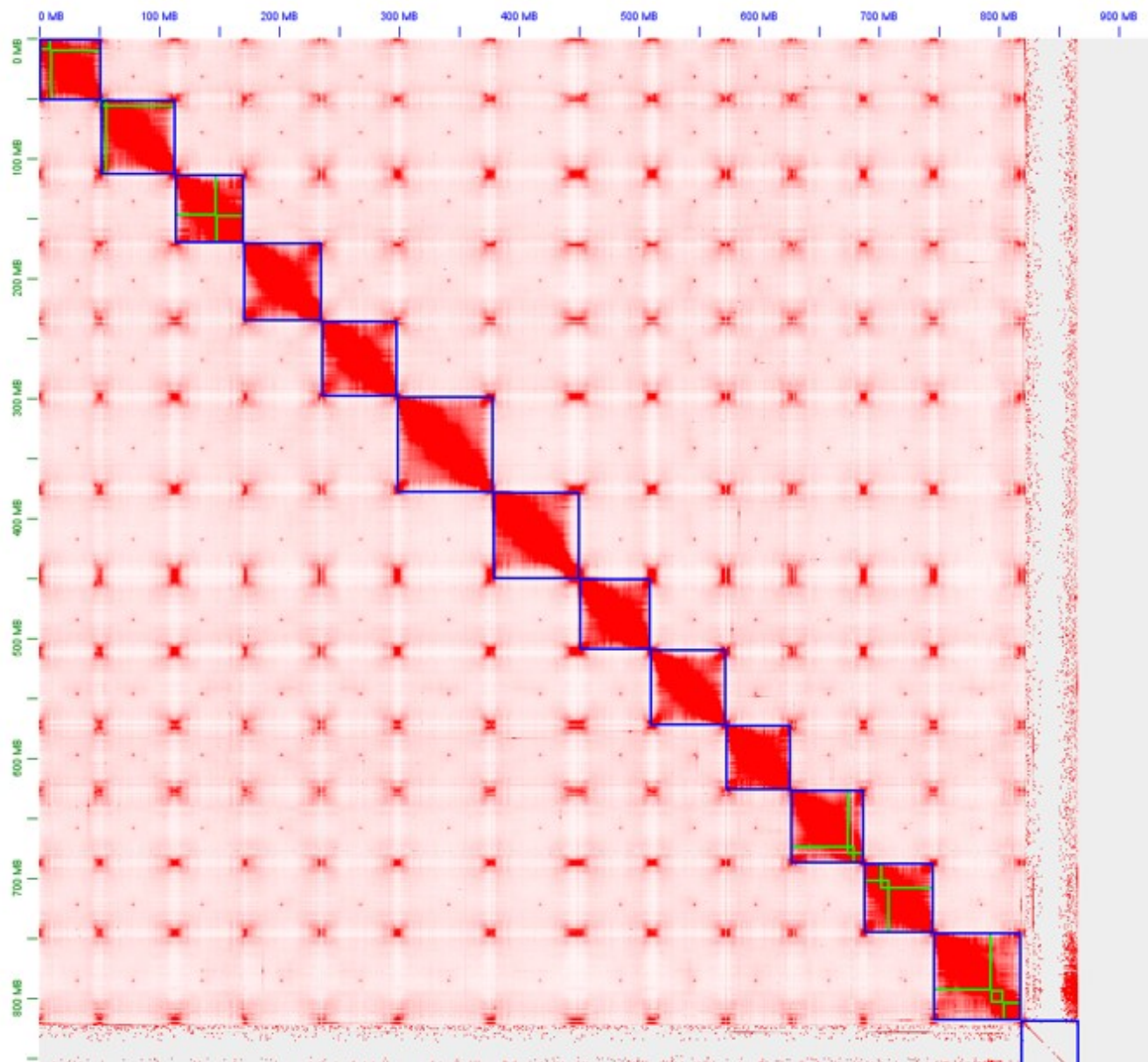
另一个调图情况，蓝圈所指的部分应该按照之前说的信号强烈的部分应该和强烈的部分挨在一起，因此我们要把中间那个contig反转一下。此外，值得注意的是，深绿色部分的HIC信号明显与周围的信号不一样比较“粗糙”，这些一般是杂合区、重复区，尤其是有的着丝粒区也会有杂合信号，因此，我们要把杂合区和杂合区放到一起，即绿色箭头所指的地方。





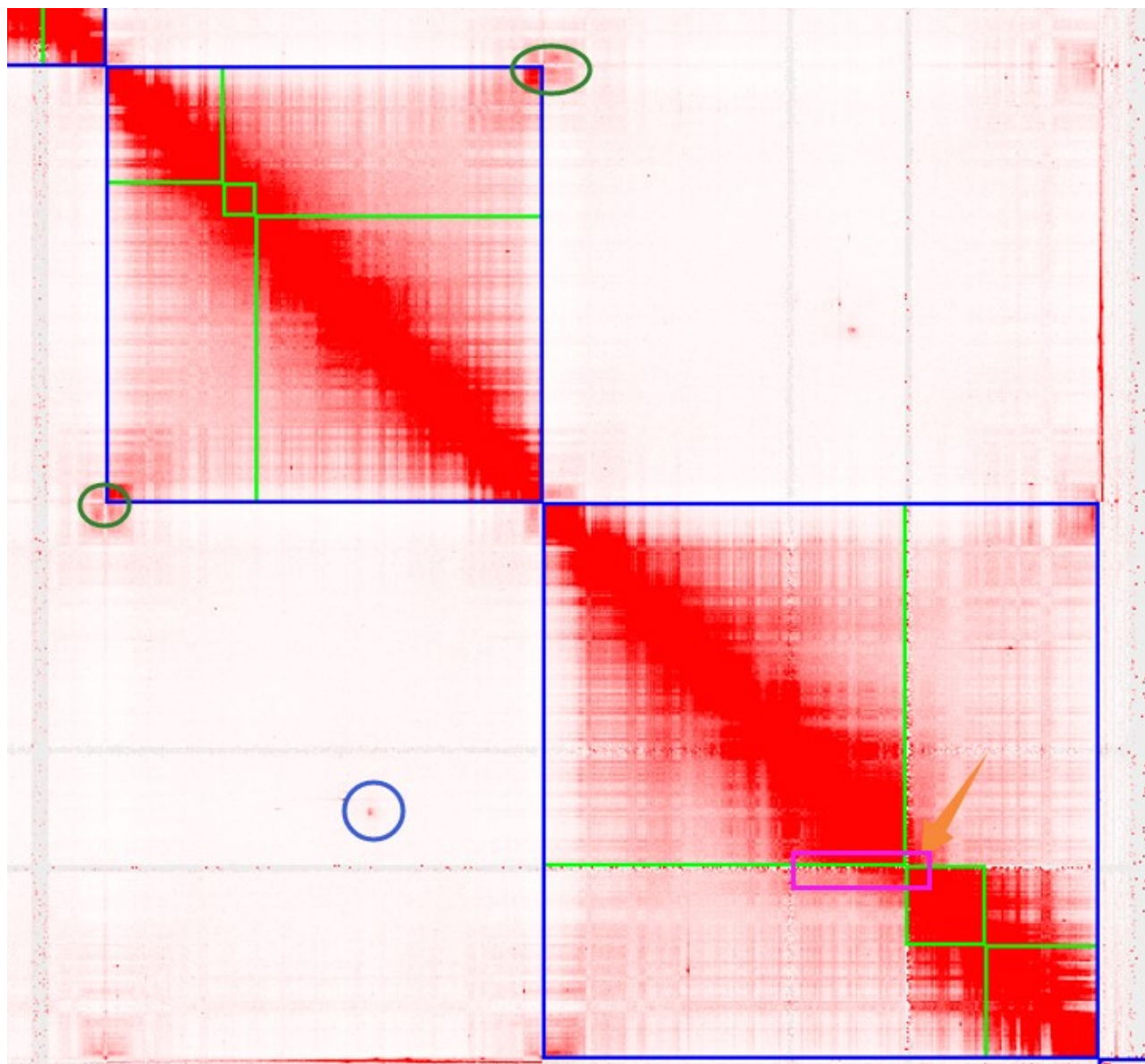
反转，之后看起来比较正常了，但对于深绿色框住的部分，这些杂合区之间的Gap，往往在填补Gap方面比较困难，实际上蓝色框住的部分与紫色框住的部分可能还有重叠区域，这种情况就是“着丝粒区杂合”，那么我们就需要根据HIC互作信号来进行删除，可能是删除紫色也可能是删除蓝色。





调完图后，整体就比较干净了。每个染色体的中间线上都有着丝粒互作（小圆点）。从整体图来看，右下角的未挂载区与最后一条大染色体是有许多互作信号的，但是由于其本身比较碎，却又不能放回去，实际可能要用ONT+HIFI+HIC的组装模式，把这个区域组装的更好更长，才能使最后一条染色体变长一些。

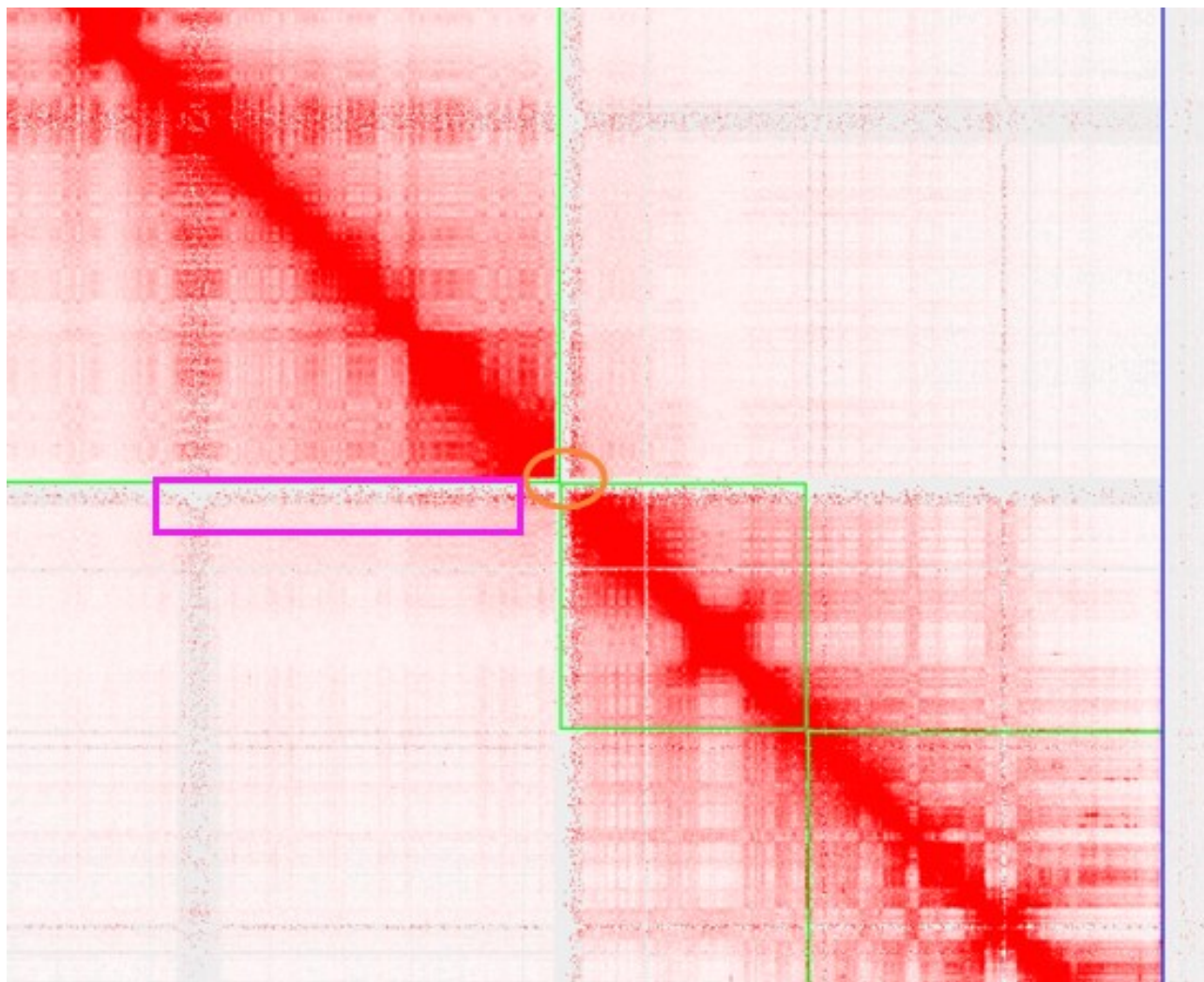




蓝圈：着丝粒互作信号。绿圈：端粒互作信号。

橙色箭头所指的位置：周围颜色没有特别强烈，不如上一条染色体整体颜色都比较粗，比较红，从紫色框柱的部分来看，确实像是缺了什么，这种一般都是复杂重复区没有组装出来。





放大的看，确实感觉少了什么，对于这些区域的组装只能：①更换组装软件；②增加测序深度和ONT超长序列；

综上，该基因组HIC调图完成。

后续：①基因组与参考基因组进行排序；②同源挂载其他contig级别基因组到当前版本基因组再做syri分析比较差异。③统计各个版本的端粒与rDNA情况，查看是否primary没组装出来的端粒在hap1/hap2的对应位置组装出来。④未挂载区进行NT比对、BUSCO评估、回比染色体部分三重比较进行过滤，提高挂载率。