

DNAcopy 分段与 test-v1→test-v2 点位 liftover 报告

科研工作台

October 13, 2026

1 输入与身份

- 分段输入: in/bins.csv (180 个 bin, chr01, 0-based 半开区间 [start, end), bin 宽 100), SHA256 = 6e66f2...67d5a9 (全值见 out 目录各 manifest)。信号模式 = log2_ratio (相对信号), 信号单位由用户声明为 synthetic relative coverage, 参考版本声明为 synthetic DNA v1。
- liftover 输入: in/points.csv(6 个点位), SHA256 = 77c676...6456a8e8; chain = in/test.chain, SHA256 = facb74...1726500 (与 in/genetic-input.json 声明一致, 已校验)。
- chain 来源与许可均为用户声明: *Synthetic exact fixture v1; chrA is UCSC target-field source, Original synthetic fixture*。该 chain 为原创合成测试 fixture(test-v1 → test-v2), 不是 hg19/hg38 或任何真实组装。

2 方法与参数

分段: DNAcopy 1.72.3 (R 4.2.2 Patched) circular binary segmentation。参数: mode=log2_ratio, permutations(nperm)=1000, seed=1729, alpha=0.01, min_width=2, undo.splits=sdundo, undo.SD=3; 未做平滑, 未做任何标准化, 未计算绝对拷贝数。坐标约定: 0-based 半开区间 [start, end); CBS 探针取 start+1。

Liftover: pyliftover 0.4.1 点位转换, 仅支持点位 (不支持区间/BED/等位基因转换)。坐标为双链上统一的 0-based 点位; 保留全部匹配, 歧义点位不自动取舍。链方向: 目标链取自 chain 目标字段 (正链块 targetStrand=+, 负链块 targetStrand=-), 反向链映射的目标坐标与正链点相同但链号随点源链翻转 (p6 例证)。

3 结果: 分段表 (相对信号)

3 个片段, 共 180 bins (chr01, 0-based 半开区间, bin 区间右端不包含):

segment	chromosome	start_0based	end_0based	bins / mean log2_ratio
1	chr01	0	6000	60 / +0.0007
2	chr01	6000	12000	60 / +1.0027
3	chr01	12000	18000	60 / +0.0009

重要: 上述为相对信号(log2_ratio)观测均值, 不代表绝对拷贝数状态; 未做纯度/倍性/GC/mappability 校正, 亦不做临床解释。检出片段即 seg2 (6000–12000) 相对信号升高约 $2^{1.0027} \approx 2.01$ 倍。

4 结果: liftover 映射与未映射

映射表 (完整, 含歧义与未映射原因; 目标 contig 保持 chrB/chr01 命名体系中的实际返回值, p5 因超出 chain 覆盖范围而无法保留 chr01/chrB 身份):

point_id	src chr	pos(0b)	strand	status	tgt chr	tgt pos(0b)	tgt strand	score
p1	chrA	110	+	unique	chrB	310	+	100
p2	chrA	160	+	ambiguous	chrB	360	+	100
p2	chrA	160	+	ambiguous	chrB	1489	-	90
p3	chrA	250	+	unmapped	—	—	—	—
p4	missing	100	+	unknown_contig	—	—	—	—
p5	chrA	1000	+	out_of_bounds	—	—	—	—
p6	chrA	160	-	ambiguous	chrB	360	-	100
p6	chrA	160	-	ambiguous	chrB	1489	+	90

4.1 未映射原因明细

- **p3 (unmapped)**: chrA:250(+) 未落入任何 chain block 覆盖区间 (块内存在空洞), 无法转换。
- **p4 (unknown_contig)**: 源 contig 名 missing 不存在于 chain 的 source 字段 (该 chain 仅含 chrA), 不参与任何匹配。
- **p5 (out_of_bounds)**: chrA:1000(+) 超出该 chain 对 chrA 的覆盖范围上界, 无法转换。
- **p2 / p6 (ambiguous)**: 点位落在两个重叠 block 的边界/重叠区, 给出两个候选 (chrB:360 与 chrB:1489), 工具保留全部匹配不自动取舍, 需人工/后续证据决定。
- **unique (p1)**: 唯一匹配, chrA:110 → chrB:310 (0-based, 双链同号)。

链方向: p6 源为反向链 (-), 其两个候选目标坐标分别出现在 chrB:360(-) 与 chrB:1489(+), 与 chain 的 targetStrand 声明一致, 反映链方向而非坐标偏移。

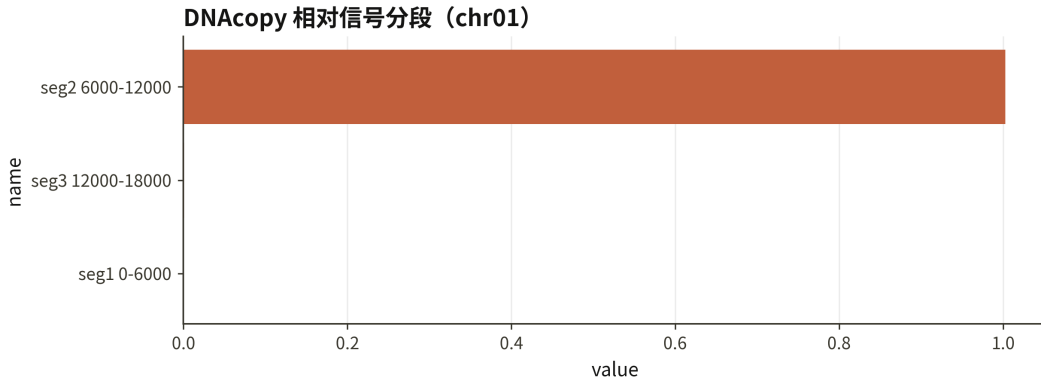


Figure 1: 三个 CBS 片段的相对信号均值 (log2_ratio; 非绝对拷贝数)。

5 限制

分段结果为观测到的相对信号分段, 非绝对拷贝数; 未做纯度/倍性/GC/mappability 校正或临床解释。chain 为用户声明的合成 fixture (test-v1 → test-v2), 文件本身不认证组装身份; liftover 仅点位、无区间或等位基因转换; 歧义点位保留全部候选不自动决策。全程保留 chr01 / chrB 等原始命名, 未重命名或映射到真实人类组装。

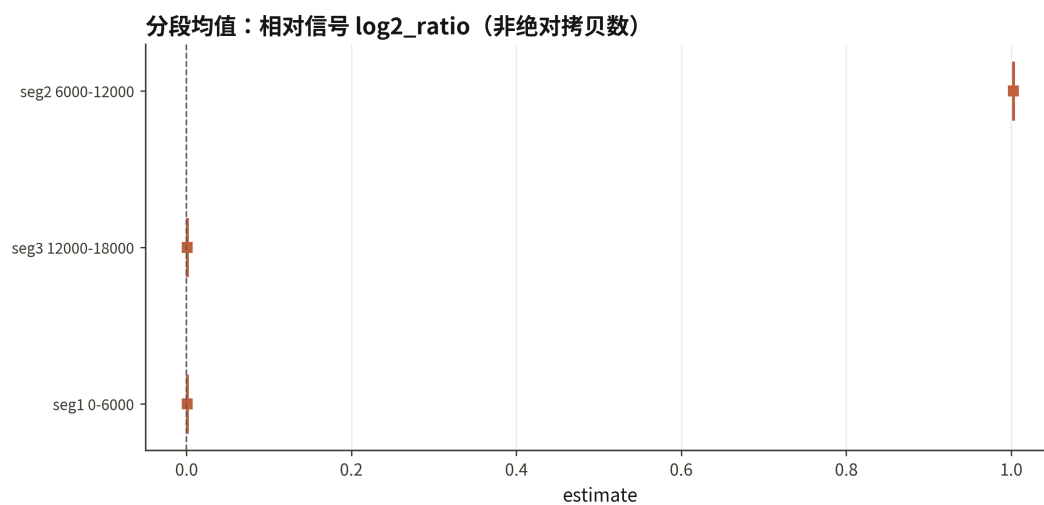


Figure 2: 片段均值森林图 (参考线 0; 横轴为相对信号 $\log_2\text{ratio}$)。

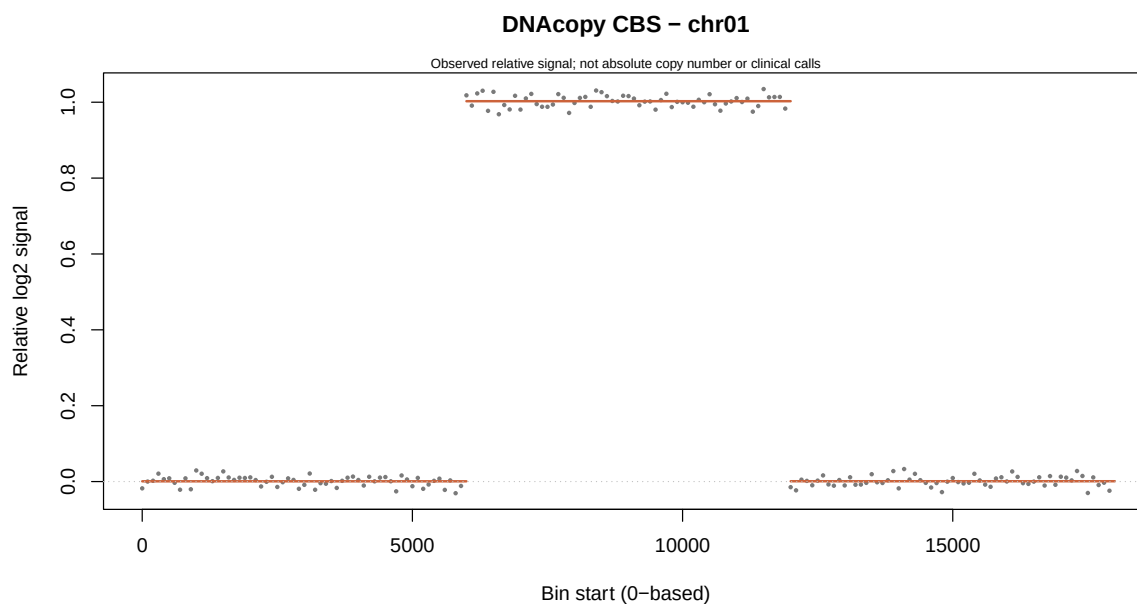


Figure 3: DNAcopy 官方诊断图 (逐 bin 信号与 CBS 分段叠加)。