

ロングリードアセンブリによる個別化参照ゲノムを用いたがんゲノム解析 パイプラインPRCGAPの開発

坂本祥駿¹, 越智陽太郎^{2,3}, 木暮泰寛⁴, 加登翔太⁵, 佐藤亜以子⁵, 須川正啓^{1,5}, 田中庸介⁶, 辻村太郎³, 三上貴司⁷, 永江玄太⁸, 千葉健一¹, 岡田愛¹, 伊藤佑¹, 鈴木創¹, 山本拓也^{3,9,10}, 油谷浩幸⁸, 古賀友紀^{11,12}, 加藤格¹³, 滝田順子¹³, 間野博行¹⁴, 小川誠司^{2,3,15}, 片岡圭亮^{4,16}, 加藤元博⁵, 白石友一¹

¹ 国立がん研究セ・研・ゲノム解析基盤開発 ² 京都大・医・腫瘍生物学 ³ 京都大・WPI-ASHBi ⁴ 国立がん研究セ・研・分子腫瘍学 ⁵ 東京大・医・小児科 ⁶ 国立がん研究セ・研・がん創薬研究 ⁷ 京都大・医・CCII ⁸ 東京大・先端研・ゲノムサイエンス ⁹ 京都大・CiRA ¹⁰ 理研AIP・iPS細胞連携医学的リスク回避 ¹¹ 九州大・小児科 ¹² 九州がんセ・小児科 ¹³ 京都大・医・小児科 ¹⁴ 国立がん研究セ・研・細胞情報学 ¹⁵ 近畿大・医 ¹⁶ 慶應大・医・血液

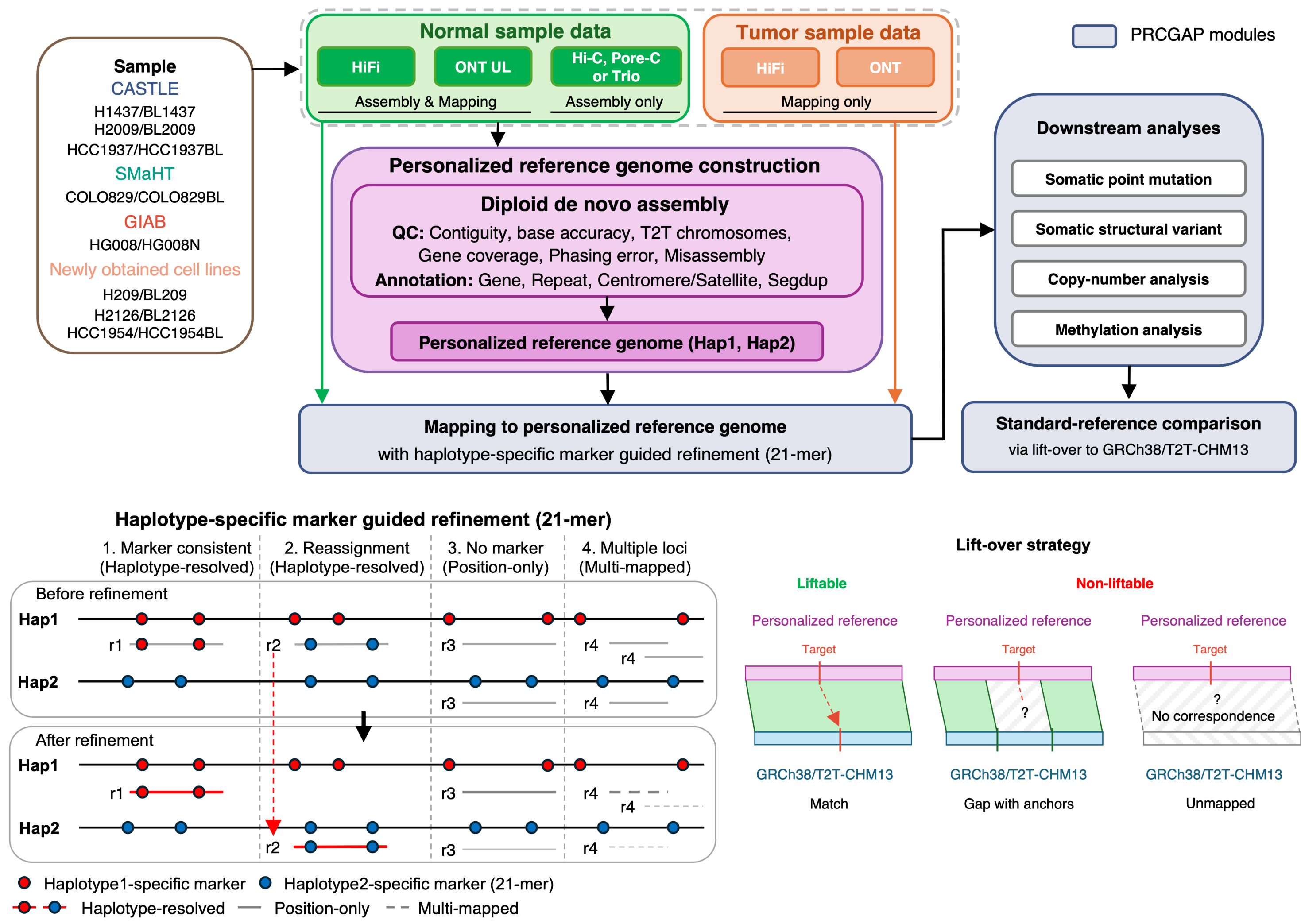
Introduction

Mutations, including point mutations and structural variants (SVs) in human cancer genomes, are typically identified by mapping sequencing data to existing human reference genomes. However, there are “dark regions” in the human reference genomes, which are highly repetitive and ethnically/individually variable regions, such as centromeres (Logsdon et al., 2024, Nature). Recent advances in long-read sequencing technologies and de novo assembly methods have facilitated the construction of nearly complete human genomes (Cheng et al., 2022, Nat. Methods; Rautiainen et al., 2023, Nat. Biotechnol.). In this study, we aimed to leverage these sequencing data and de novo assembly methods to comprehensively analyze cancer genomes, including highly repetitive and ethnically/individually variable regions, by using personalized reference genomes.

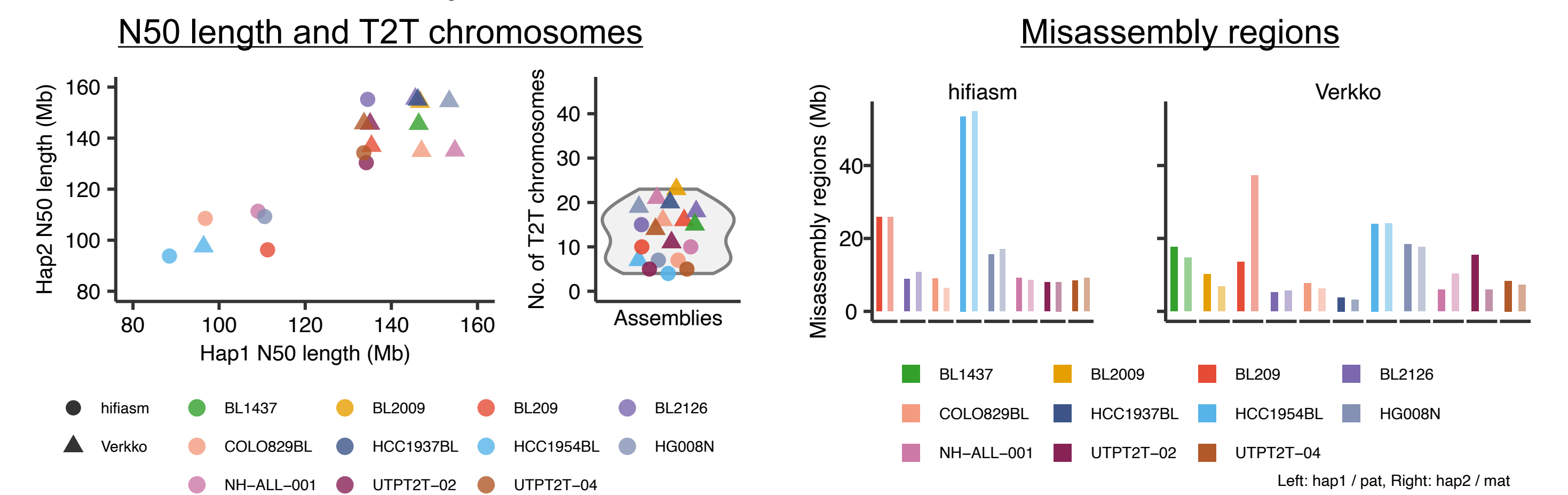
We developed **Personalized Reference genome-based Cancer Genome Analysis Pipeline (PRCGAP)** to perform cancer genome analyses based on personalized reference genomes constructed by de novo assemblies with Oxford Nanopore ultra-long reads, PacBio Hi-Fi, and Hi-C/Pore-C data. We tried to apply PRCGAP to eight tumor-normal cell-line pairs (five pairs: publicly available, three pairs: newly generated).



Personalized Reference genome-based Cancer Genome Analysis Pipeline (PRCGAP)

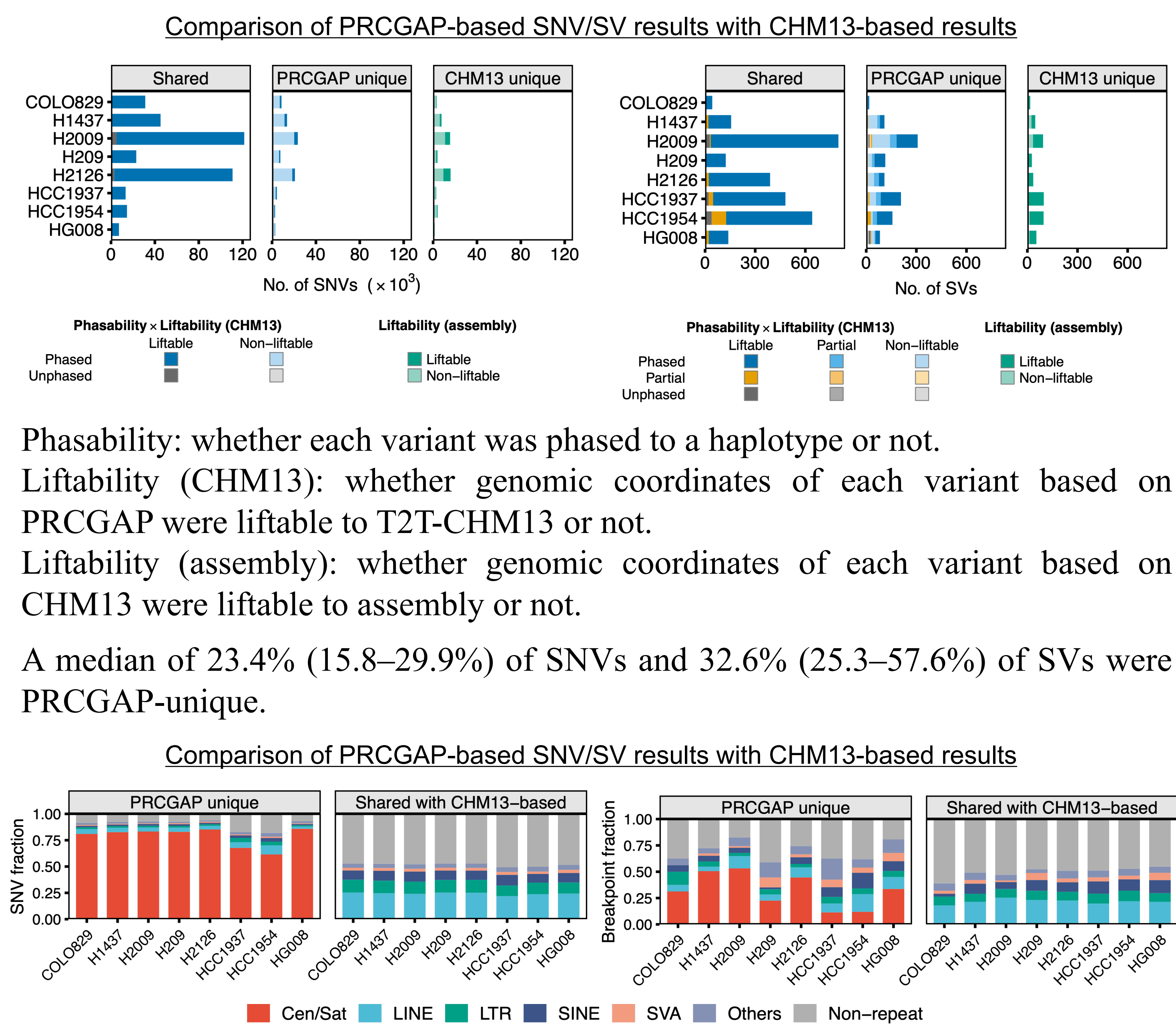


De novo assembly results



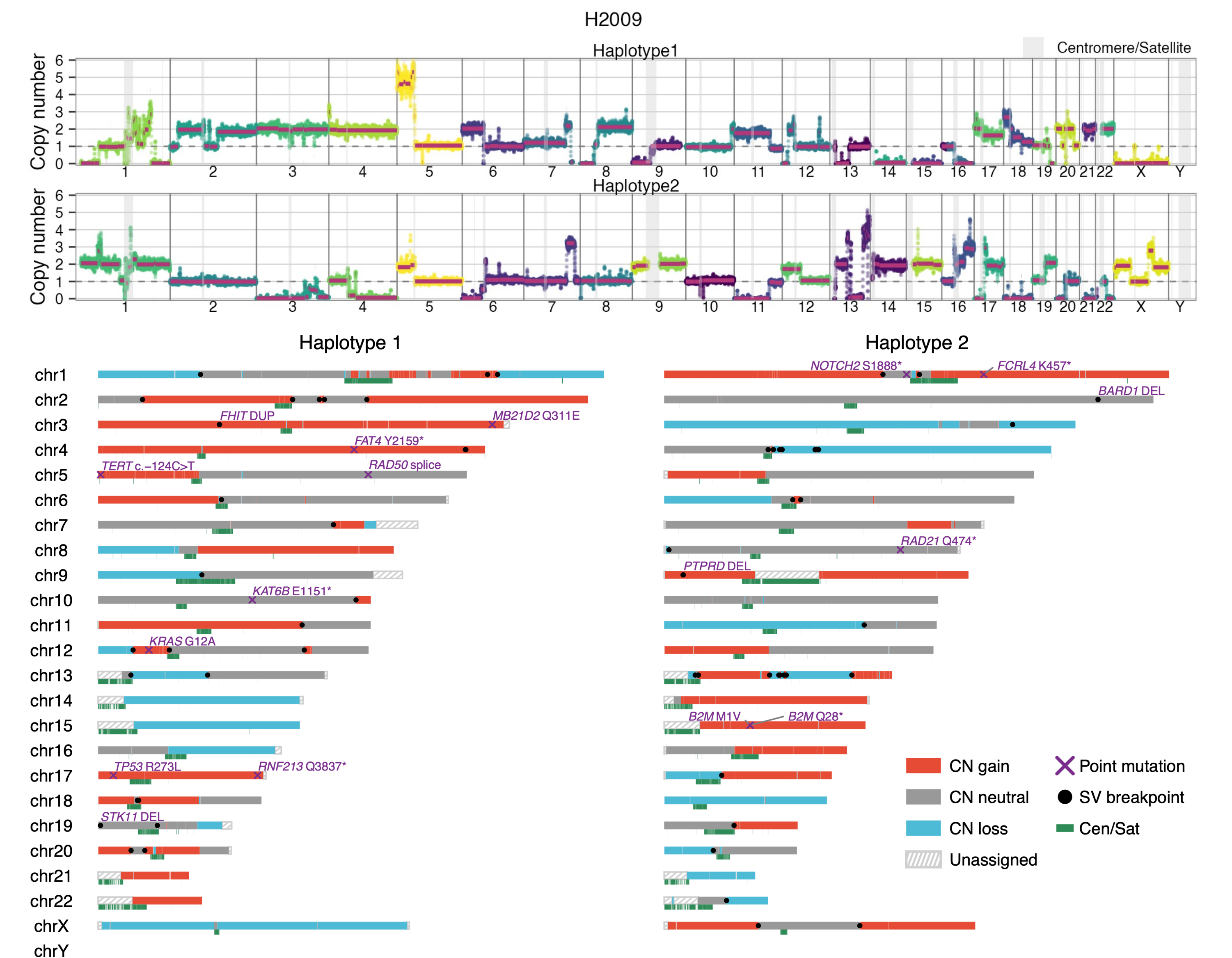
High quality de novo assemblies were obtained with hifiasm and Verkko.

Benchmark of somatic variants

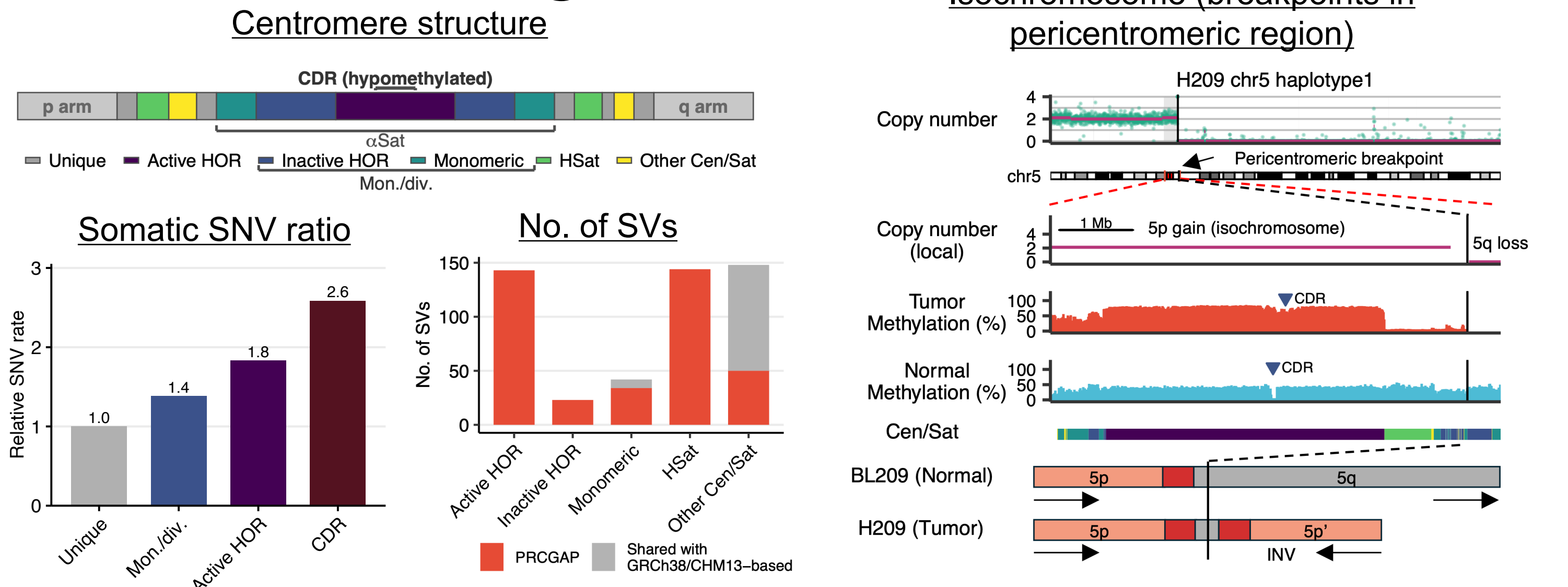


PRCGAP-unique variants were enriched in regions annotated as centromeres.

Copy number analysis and genomic overview

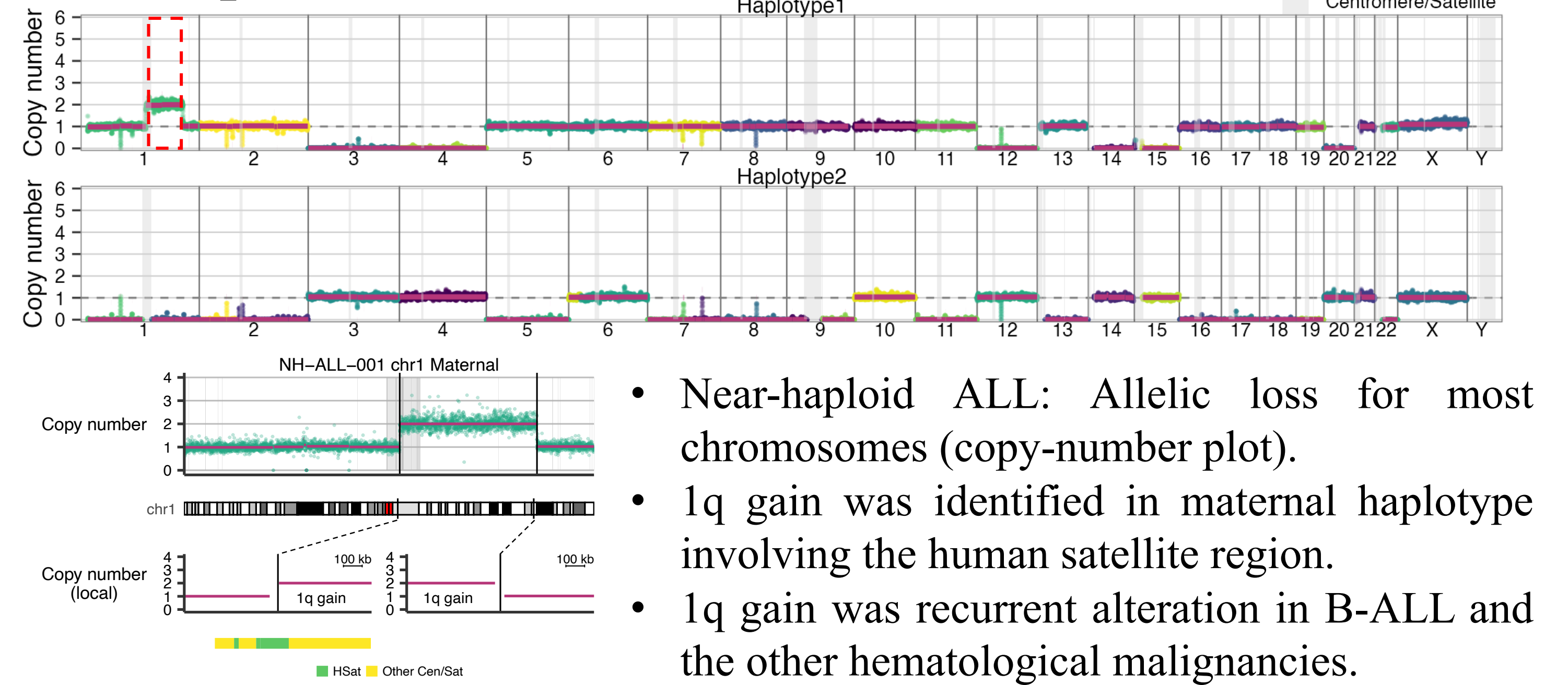


Centromere-involving variants



- Somatic SNVs were enriched in centromere dip regions (CDRs), the putative kinetochore formation sites.
- Most of the SVs in α -satellite and human satellite regions were undetectable by GRCh38- or T2T-CHM13-based pipelines.
- Analysis of the H209 cell line revealed a pericentromeric breakpoint associated with isochromosome 5p in chromosome 5 haplotype 1, and the resulting isochromosome retained the CDR.

Near-haploid ALL



Discussion

Personalized reference genome-based analysis is expected to expand into clinical settings (Sabbagh et al., Nature Genetics 2026; Ashley et al., Nature 2026). PRCGAP provides a comprehensive framework that anticipates this shift.