

RESEARCH

Open Access



Pervasive chromosomal instability drives the karyotypic evolution of hypodiploid tumours

Elle Loughran^{1,2*}, Aoife McLysaght¹ and Máire Ní Leathlobhair^{1,2*}

Abstract

Background Tumours frequently exhibit extreme levels of aneuploidy. While increases in ploidy are well-characterised, the opposite phenomenon—extensive chromosome loss leading to hypodiploidy—remains underexplored.

Methods Here, we analyse over 17,000 cancer genomes from 34 cancer types and perform a pan-cancer analysis of karyotypic evolution in hypodiploid tumours. We develop methods to identify current and former hypodiploid tumours, analyse predictors of chromosome loss patterns across tissues, and characterise the relationship between hypodiploidy and other forms of chromosomal instability.

Results We find that hypodiploidy is widespread and associated with a generalised chromosomal instability phenotype, marked by significantly elevated rates of genome doubling, intrachromosomal copy number alterations, chromoanagenesis, and intra-tumour heterogeneity. These tumours are hypoxic and strongly enriched for *TP53* mutations. However, we also identify a subset of cancers—acute lymphoblastic leukaemia (ALL), kidney chromophobe, and adrenocortical carcinoma—that exhibit stable hypodiploidy, with stereotyped chromosome loss patterns, low chromosomal instability, and distinct evolutionary origins. We exploit this stability to develop a simple method of distinguishing poor-prognosis masked hypodiploid from good-prognosis hyperdiploid ALL using only cytogenetic data, enabling more precise risk stratification. Finally, we show that unstable hypodiploidy predicts poor prognosis across cancers. Genome doubling does not confer a fitness advantage in hypodiploid tumours, nor do these tumours evolve to avoid loss of dosage-sensitive genes.

Conclusions Together, these findings provide the first pan-cancer characterization of hypodiploidy as a widespread and clinically relevant phenomenon often driven by pervasive chromosomal instability, and illustrate the remarkable ability of cancer cells to tolerate and evolve under extreme dosage imbalance.

Keywords Hypodiploidy, Aneuploidy, Chromosomal instability, Genome doubling, Karyotype evolution, Chromosome loss, Masked Hypodiploidy, Acute lymphoblastic leukaemia, Tumour heterogeneity

*Correspondence:

Elle Loughran
loughrae@tcd.ie
Máire Ní Leathlobhair
nleathlm@tcd.ie

¹Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland

²Moyne Institute of Preventive Medicine, Trinity College Dublin, Dublin 2, Ireland



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

Aneuploidy is a hallmark of cancer, observed in 90% of solid tumours and a majority of haematological cancers [1, 2]. The extent of aneuploidy varies significantly, both within and between cancer types: while many tumours possess one or a few recurrent aneuploidies, often affecting chromosomes carrying specific cancer driver genes [3, 4], others exhibit widespread aneuploidy across much of the genome, resulting in marked deviations from diploidy. Extreme aneuploidy often results from whole genome duplication (WGD), which is the second most common alteration in cancer and is associated with chromosomal instability, tumour progression, and poor prognosis [5–7].

More rarely, tumours may lose large numbers of chromosomes to reach a hypodiploid karyotype. This phenomenon has been best studied in acute lymphoblastic leukaemia (ALL), where near-haploid (<30 chromosomes) and low-hypodiploid (30–39 chromosomes) cases form a rare subgroup with distinct mutational profiles [8], poor treatment response [9, 10] and very poor prognosis [11]. Near-haploid or low-hypodiploid subtypes have also been observed in other cancers including kidney chromophobe carcinoma [12], chondrosarcoma [13], giant cell glioblastoma [14], malignant pleural mesothelioma [15] and oncocytic follicular thyroid carcinoma [16]. In at least some tumour types, hypodiploidy can arise early in tumour development [17] and persist throughout metastatic progression [18].

Hypodiploid tumours frequently undergo WGD [13, 19], resulting in hyperdiploid or near-triploid karyotypes with homozygous disomies and heterozygous tetrasomies. It has been proposed that WGD is positively selected in hypodiploid tumours to normalize gene dosage [11], and/or buffer the effects of deleterious mutations in regions with hemizygous deletions or loss of heterozygosity [20]. In ALL, WGD is rare overall but occurs in 64% of near-haploid and 44% of low-hypodiploid cases [11]. In some patients, genome-doubled and non-doubled subclones can coexist, with the doubled subclone often dominating at diagnosis and the hypodiploid founder clone at relapse [1]. Because the doubled clone retains the poor prognosis of its hypodiploid precursor but may be misclassified as a good-prognosis high-hyperdiploid case ('masked hypodiploidy'), WGD presents an important diagnostic challenge in hypodiploid ALL [21, 22].

Hypodiploid subtypes in cancer types including ALL [21], kidney chromophobe cancer [23] and Hürthle cell carcinoma [18] exhibit tissue-specific stereotyped aneuploidy patterns, with recurrent losses of specific chromosomes. The biological basis for these patterns is generally unknown, although it has been suggested that chromosome 7 may be constrained against monosomy in thyroid

cancer cells due to the cancer-essential imprinted genes it carries [24], and that chromosome 21 is consistently retained in hypodiploid ALL due to its leukemogenic effects [25]. The mechanisms by which these karyotypes arise are also unclear, but could involve successive losses or multipolar mitosis [1]. Beyond its prognostic significance, an improved understanding of hypodiploidy thus offers insight into the cellular constraints and selective forces that shape tumour karyotypes in the context of extensive chromosome loss.

In this work, we perform a pan-cancer analysis of hypodiploidy in order to understand the prevalence, origins, and evolution of low-hypodiploid and near-haploid tumours, particularly outside ALL. Using data from 17,239 tumours in The Cancer Genome Atlas (TCGA) and the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer [26, 27], we identify hypodiploid tumours across a wide range of haematological and solid cancers and analyse patterns of chromosome loss with respect to tissue type and chromosome features. We characterise the relationship between hypodiploidy and chromosomal instability within and between cancer types, and identify genomic and environmental factors associated with hypodiploidy. This work distinguishes classes of stable and unstable hypodiploid tumours, maps an underlying chromosomal instability phenotype that connects copy number alterations at multiple scales, and contributes to our knowledge of the origins of extreme aneuploidy.

Methods

Cohort Descriptions

This study utilised publicly available cancer genomics datasets from The Cancer Genome Atlas (TCGA), the Pan-Cancer Analysis of Whole Genomes (PCAWG) consortium, and the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. All samples analysed in this study were collected and processed by the respective consortia under previously described protocols, including ethical approval and informed consent.

TCGA cohort

We analysed publicly available TCGA data comprising 10,332 tumour samples across 33 cancer types. Analyses were restricted to primary tumours with available copy number variation (CNV) data. Specifically, we used segment-level ASCAT data, MAF files, and RNA-seq count tables from GDC. FFPE samples were removed. Where multiple samples were available per patient, one sample per patient was retained. Data on pathogenic germline mutations identified in the TCGA cohort by Huang et al [28] was obtained from GDC. Tumour types were defined according to the TCGA study classification scheme:

<https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations>.

TARGET-ALL cohort

We downloaded ASCAT2-derived allele-specific copy number profiles for all TARGET-ALL-P2 cases ($N = 293$) from GDC using TCGAbiolinks, retaining one sample per patient.

PCAWG cohort

We analysed data from the PCAWG study comprising whole genome-sequenced data from 1,699 donors across 20 cancer types. Analyses focused on whitelisted samples with consensus copy number and WGD calls that were not included in the TCGA project. PCAWG data is available from <https://object.genomeinformatics.org> via the AWS S3 interface.

Mitelman database

Data were obtained from the Mitelman Database, a curated resource of chromosomal aberrations and gene fusions derived from published studies. We extracted karyotypes corresponding to “acute lymphoblastic leukaemia/lymphoblastic lymphoma” and analysed all samples for which CytoConverter successfully identified copy number aberrations.

Copy number data

We downloaded all acute lymphoblastic leukaemia karyotypes from the Mitelman Database of Chromosome Aberrations & Gene Fusions in Cancer and used CytoConverter [26], bedops partition [29] and bedtools map [30] to calculate net copy number alterations for each genomic region. We removed cases with errors from CytoConverter or with unknown chromosome numbers, then calculated copy number as $2 + \text{net copy number alteration}$. Cases with 46 chromosomes and no numerical abnormalities were inserted separately as diploids. For the TCGA tumours, allele-specific copy number profiles computed using ASCAT [31] were downloaded from the Genomic Data Commons using TCGAbiolinks [32]; we removed metastatic or recurrent tumours, FFPE samples and ASCAT-defined suspected contaminated samples, and deduplicated to keep one sample per patient. We ran MEDICC2 [33] on ASCAT allele-specific absolute copy number profiles for all TCGA samples and defined a sample as WGD-positive if the minimum event distance was decreased when WGD was included as a possible evolutionary step. Chromosome somy was calculated as the mean copy number along the chromosome, rounded to a whole number. Samples with nullisomic chromosomes or sub-haploid autosome counts were excluded. Autosome counts were used instead of chromosome counts due to

the unreliability of sex chromosome copy number assignments in these datasets.

Identifying hypodiploids

The threshold for hypodiploidy was set at ≥ 6 autosome losses, which corresponds to ≤ 38 autosomes. This threshold was chosen based on the distributions of autosome counts in two cancer types with established hypodiploid subtypes (ALL and KICH). Samples with 23–27 autosomes were classified as near-haploid, while samples with 28–38 autosomes were classified as low-hypodiploid. Tumours with 39–43 autosomes were considered high-hypodiploid and grouped with the less-extreme aneuploid samples in this analysis. Samples with 49–65 autosomes were classified as hyperdiploid based on the 51–67 chromosome range defined for ALL [34], subtracting two due to the lack of sex chromosomes.

In the TCGA dataset, chromosome losses were ascertained using LOH information, which allowed us to identify former hypodiploids that had since undergone WGD or otherwise gained chromosomes. Autosomes with LOH (i.e. a minor copy number of 0 in the ASCAT copy number profile) over $>90\%$ of their length were defined as ‘lost’ and assigned a $\text{somy}_{\min}$ of 1, while other autosomes were assigned a $\text{somy}_{\min}$ of 2. Minimal autosome count ($\text{autosome}_{\min}$) during the tumour’s evolution was computed as the sum of $\text{somy}_{\min}$ across all autosomes. While individual LOH events can occur after WGD if the same homolog is lost twice, meaning that the tumour was never necessarily in a hypodiploid state, we assume that LOH events occurred pre-WGD and count tumours with an $\text{autosome}_{\min} \leq 38$ as (former) hypodiploids. Where LOH information was unavailable, i.e. in the Mitelman-ALL dataset, hypodiploids were identified based on current autosome count at time of karyotyping, calculated as the sum of rounded chromosome ploidies.

We classified all TCGA genomes into one of five groups based on their ASCAT copy number profiles: near-haploid ($< 28 \text{ autosome}_{\min}$), low-hypodiploid ($\leq 38 \text{ autosome}_{\min}$), diploid (all autosomes disomic), polyploid (positive WGD call and no hypodiploid history), and aneuploid (all other cases, i.e. non-diploid but without extreme enough copy number alterations to qualify as polyploid, low-hypodiploid or near-haploid).

Distinguishing masked hypodiploids from hyperdiploids using cytogenetic data

We calculated a masked hypodiploidy (MH) score for each sample as $\# \text{ tetrasomic chromosomes} - \# \text{ trisomic chromosomes}$. To test the ability of the score to distinguish masked hypodiploids from hyperdiploids, we gathered two datasets of masked hypodiploids (positives) and hyperdiploids (negatives). For the positives, we took Mitelman ALL cases that had both a hypodiploid and a

non-hypodiploid clone, built copy number phylogenetic trees with MEDICC2 and selected cases where WGD was inferred and the non-hypodiploid clone had at most 76 autosomes ($N = 123$). For the negatives, we obtained allele-specific hyperdiploid copy number data from Woodward et al. (2023) ($N = 577$) and counted chromosomes labelled as “Tri” vs “Tetra_2_2” or “Tetra_3_1” after excluding the sex chromosomes and chromosomes with subclonal copy number variation. We additionally used the TARGET-ALL-P2 dataset ($N = 293$, 292 non-hypodiploid) to assess accuracy outside the high-hyperdiploid range. ASCAT allele-specific copy numbers were downloaded, deduplicated per patient and hypodiploidy inferred as described for the TCGA dataset.

Chromosome loss patterns

We created a binary loss/retention profile for each chromosome arm in each sample, defining a chromosome arm as lost if it had >90% LOH (TCGA samples) or had copy number < 2 for >90% of its length (Mitelman samples). The sex chromosomes and acrocentric arms (13p, 14p, 15p, 21p, 22p) were excluded due to lack of data. The average loss rate for each chromosome arm was calculated separately by tissue. We clustered all samples by their chromosome loss profiles using the Jaccard distance ($1 - \text{number of shared lost chromosome arms} / \text{total number of lost chromosome arms between the two samples}$), and calculated the median Jaccard similarity ($1 - D$) within and between each cancer type. The heatmap was created with ComplexHeatmap [35] in R. To investigate factors associated with chromosome loss rate, we obtained chromosome tumour suppressor - oncogene density scores from Davoli et al. (2013), chromosome distances from the nuclear surface (for all chromosomes except 9, 22 and Y) from Girelli et al. (2020) [36], chromosome lengths from the D3GB R package [37], and gene haploinsufficiency probabilities from Collins et al. (2022) [38]. These haploinsufficiency probabilities were summed for each chromosome to create a chromosomal dosage sensitivity score. We performed linear regression of chromosome loss rate on chromosome length, distance from the nuclear surface, driver gene density and dosage sensitivity score pan-cancer and separately by tissue using glm() in R. To assess whether the overall degree of loss affects these relationships, we re-ran the pan-cancer regression analysis for the TCGA tumours after stratifying samples by their number of chromosome losses (1, 2, 3, 4, 5, 6+, where 6+ indicates current or former hypodiploid samples). Cancer types were defined as stereotyped hypodiploids if the median Jaccard similarity between low-hypodiploid chromosome arm loss profiles from tumours of that cancer type was higher than all median similarities between that cancer type and other

cancer types. For the ALL cases, only current hypodiploid cases (autosome count ≤ 38) were included.

Chromosomal instability

TCGA cases were defined as WGD-positive if they had a positive MEDICC2 WGD call. ALL cases were defined as WGD-positive if they had ploidy > 2.7, if they had a hypodiploid and non-hypodiploid clone and a positive MEDICC2 WGD call, or if they had an MH score > 0. Fully-masked clones (MH score > 0, no current-hypodiploid subclone) were classified as former near-haploids or former low-hypodiploids based on their autosome count divided by two. For the MH score criterion, we initially included all inferred fully-masked clones; in a supplementary analysis, clones with MH score > 0 but autosome count outside the well-characterised hyperdiploid range were excluded as masked hypodiploids and reclassified as non-hypodiploid and non-doubled (unless eligible under the ploidy criterion). Cases with any WGD-positive clones were counted as WGD-positive.

Intrachromosomal instability was measured for each TCGA sample by counting the number of copy number segments in each sample, and by computing for each chromosome the proportion of the summed length of all segments occupied by its longest contiguous segment. Chromoanagenesis status for each TCGA sample was derived from Rasnic & Linial (2021) [39] and chromothripsis status (number of ‘high confidence’ or ‘linked to high confidence’ chromothriptic chromosomes) was derived from Cortes-Ciriano et al. (2020) [40]. Copy number heterogeneity values were obtained from van Dijk et al. (2021). When computing the correlation between WGD rate and hypodiploidy rate across cancer types, hypodiploid samples were excluded from the set of samples used to calculate WGD rate to disentangle the increased WGD rate in hypodiploid tumours from the hypothesised underlying chromosomal instability phenotype driving both WGD and hypodiploidy. Similarly, median segment counts and median copy number heterogeneity were calculated based on samples without gross ploidy alterations (hypodiploidy or polyploidy). Two cancer types, ACC and KICH, were excluded from these correlations due to their extremely high rates of hypodiploidy. When controlling for level of aneuploidy in the chromosomal instability analyses, we quantified ‘level of aneuploidy’ based on the number of disomic chromosomes in the sample, with more aneuploid samples having fewer disomic chromosomes.

Survival analysis

Survival analysis was performed using Cox proportional hazards regression with the R survival [41] and survminer [42] packages as follows:

overall_survival ~ ploidy class + age at index + gender + race + cancer type

progression-free survival ~ ploidy class + age at index + gender + race + cancer type

Tumour stage was included as a covariate in specific analyses; we grouped substages from the variable 'ajcc_pathologic_tumor_stage' into Stages I, II, III or IV (e.g. Stage IIIA → Stage III) and removed cases with a stage that was missing or could not be classified into one of these four groups. We obtained survival data (overall survival and progression-free survival status and time) from the TCGA-CDR paper [43]. For PCAWG, overall survival was obtained from the August 2016 (version 9) donor clinical dataset.

Genomic correlates of hypodiploidy

We downloaded MAFs for each TCGA cancer type from the Genomic Data Commons and calculated the total number of mutations, the ploidy-corrected mutation rate (mutations/ploidy) and the number of non-synonymous mutations per sample. To detect genes enriched for mutations in low-hypodiploid samples compared to diploid samples, we performed logistic regression controlling for total non-synonymous mutation count and cancer type, using only genes present in the COSMIC Cancer Gene Census and with ≥ 5 total non-synonymously mutated samples across the low-hypodiploid and diploid cohorts. This analysis method was repeated to compare polyploids to diploids, low-hypodiploids to aneuploids, and low-hypodiploids to near-haploid cases. Purity estimates were obtained from the consensus purity estimates (CPE) calculated by Aran et al. (2015) [44] based on ESTIMATE, ABSOLUTE, LUMP and immunohistochemistry purity values.

mutated ~ ploidy_class + total_non_syn_mutations + purity + cancer_type

To analyse germline mutations, we performed a similar analysis using a list of prioritised pathogenic germline variants from Huang et al. (2018). We filtered for genes in which at least 5 patients with tumours of the relevant ploidy classes carried a pathogenic germline mutation, and performed logistic regression as follows:

mutated ~ ploidy_class (diploid or low-hypodiploid) + cancer_type

Mutational signature activities were estimated using SigProfilerAssignment v0.1.8 [45] in exome mode based on mutational spectra computed for TCGA exome sequences by the PCAWG project. Signatures were assigned based on COSMIC 3.4 SBS96 [46].

RNA-seq counts were downloaded from the Genomic Data Commons and normalised using cpm() in edgeR. We computed hypoxia scores for every TCGA sample using the Ragnum pimonidazole signature [47] as described by [48]: for each gene in the Ragnum signature,

a sample was assigned a score of +1 if its expression for that gene was higher than the median expression of that gene across all samples and -1 otherwise, and scores were summed for each sample. To confirm that our results were robust to choice of hypoxia signature, we additionally analysed the relationship between hypodiploidy rate and hypoxia score using eight hypoxia scores (Ragnum, Buffa, Winter, West, Sorensen, Elvidge, Hu and Seigneuric2) precomputed by Bhandari et al. (2019) for a subset of 20 cancer types. We computed correlations between hypoxia scores from each signature and both hypodiploidy and polyploidy. We also computed adjusted versions of these correlations by removing hypodiploid and polyploid samples (respectively) from the set of samples used to compute tissue median hypoxia scores. Microsatellite instability scores (MANTIS scores) were obtained from Bonneville et al. (2017) [49], and samples were classified as MSI-H if their MANTIS score was > 0.4 . The proliferative index for each sample was calculated as the median expression (CPM) of 117 genes included in the metaPCNA proliferation signature. This signature was derived by Venet et al. (2011) [50] from the top 1% of genes correlated with the proliferation nuclear cell antigen across healthy samples.

Statistical analysis

Multiple testing correction was performed by calculating the Benjamini-Hochberg False Discovery Rate; tests with $FDR < 0.05$ were considered significant. Correlations were computed using the Pearson correlation coefficient. The analyses in this paper were performed using R (v4.2), data.table v1.16.2 [51], tidyverse 2.0.0 [52], janitor 2.2.0 [53], TCGAbiolinks 2.34.0, TCGAutils 1.26.0 [54], openxlsx 4.2.7.1 [55], ggpubr 0.6.0 [56], ggrepel 0.9.6 [57], patchwork 1.3.0 [58], ComplexHeatmap 2.22.0, ggplotify 0.1.2 [59], RColorBrewer 1.1.3 [60], survival 3.7.0, survminer 0.5.0, broom 1.0.7 [61], DescTools 0.99.58 [62], D3GB 2.0, figpatch 0.3.0 [63] and MicroViz 0.12.7 [64]. The schematic in Fig. 2a was created using BioRender.

Results

Hypodiploidy is widespread across cancer types

To understand patterns of hypodiploidy pan-cancer, we began by identifying hypodiploid cases across 33 cancers from TCGA and comparing their chromosome count distributions to those observed in hypodiploid ALL. We first established a reference distribution for hypodiploid chromosome content by analyzing 7,922 clones from 6,907 karyotyped ALL cases in the Mitelman Database. Using CytoConverter, we inferred copy number profiles from karyotype strings and visualised the distribution of autosome counts in ALL cases with fewer than 44 autosomes (Fig. 1a). As well as a large 43-autosome peak corresponding to near-diploid ALL [1], we observed

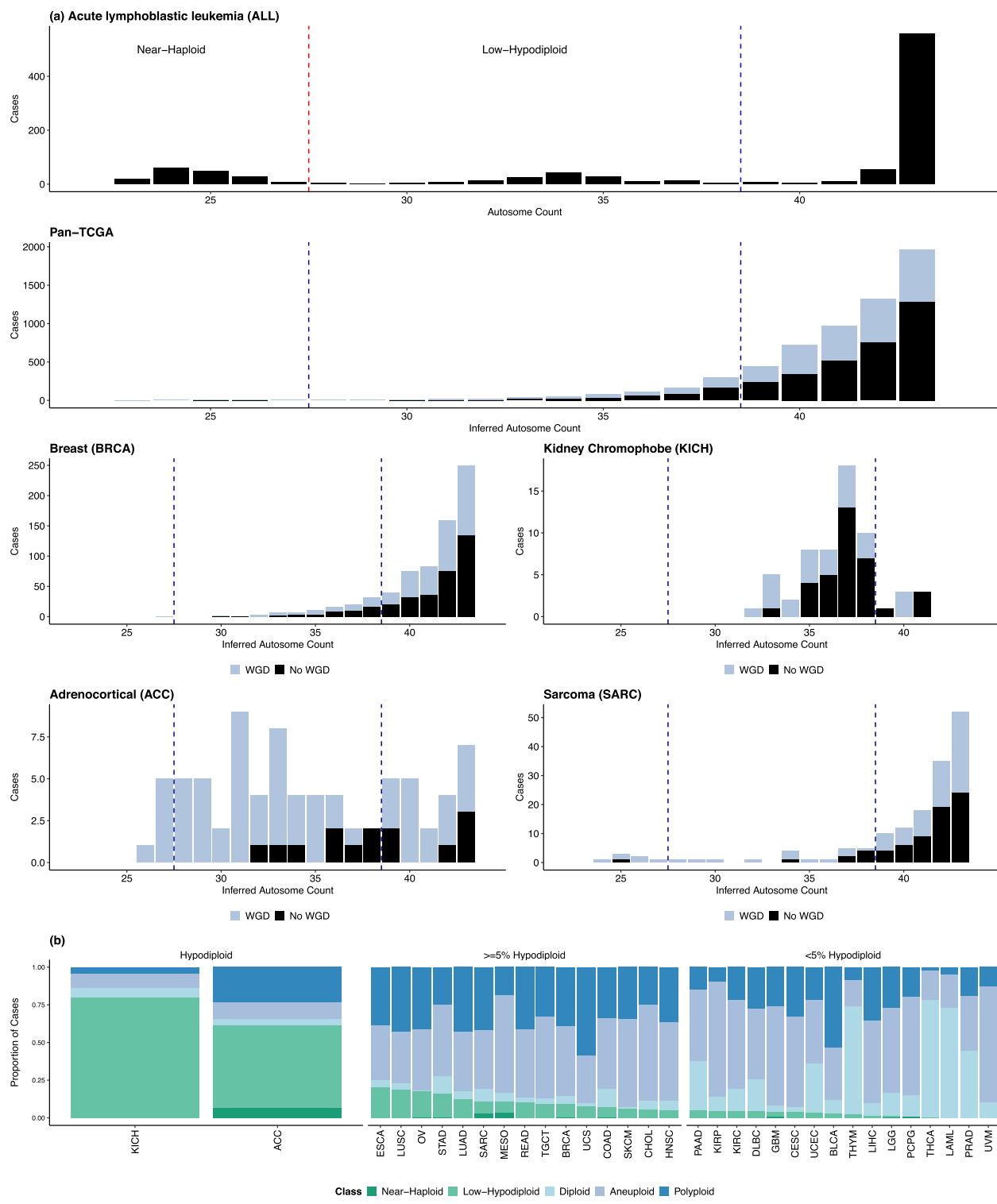


Fig. 1 Prevalence of hypodiploidy in the TCGA and Mitelman datasets. **A** Inferred sub-diploid (< 46 chromosomes, < 44 autosomes) autosome count distributions for selected cancer types. Minimal autosome counts were inferred in the TCGA dataset based on loss of heterozygosity to identify former hypodiploids which have since gained chromosomes or undergone WGD. The ALL plot is based on cytogenetic data and shows only cases with < 44 autosomes at sequencing, which are assumed not to have undergone WGD. Dashed lines indicate our thresholds for low-hypodiploidy and near-haploidy. **B** Proportions of each TCGA cancer type that are near-haploid (< 28 autosomes), low-hypodiploid ($\leq$ 38 autosomes), diploid (all autosomes disomic), polyploid (positive WGD call and no hypodiploid history), and aneuploid (all other cases). Current and former hypodiploids are both counted as hypodiploid. Absolute case counts are shown in Additional File 1: Fig. S1. Tumour abbreviations are reported as per <https://gdc.cancer.gov/resources-tcga-use/rs/tcga-code-tables/tcga-study-abbreviations>

a low-hypodiploid peak with 32–37 autosomes (129 clones) and a near-haploid peak with 23–27 autosomes (165 clones). Only 15 clones fell into the 28–31 autosome range between the two peaks. This distribution recapitulates the known bimodal distribution of chromosome counts in hypodiploid ALL [11]. The lowest autosome count was 23, observed in 18 near-haploid cases with disomy 21.

We additionally examined the distribution of autosome counts in kidney chromophobe cancer within the TCGA cohort, which has an established hypodiploid subtype [65, 66], and found that sub-diploid cases clustered around 35–38 autosomes. We used the distributions of autosome counts in ALL and KICH to set the threshold for hypodiploidy in this study at ≤ 38 autosomes, which corresponds to ≥ 6 autosome losses. This threshold resulted in the identification of 175 TCGA cases (1.7%) as hypodiploid at time of sequencing. To account for cases where hypodiploidy may have been masked by genome doubling or subsequent chromosome gains, we also identified an additional 646 cases (6.3%) with loss of heterozygosity (LOH) across ≥ 6 autosomes but autosome counts above 38, a genomic signature consistent with past hypodiploidy.

In contrast to the bimodal distribution observed in hypodiploid ALL, autosome losses in the TCGA cohort were exponentially distributed, with karyotype frequency decreasing with the degree of genomic loss. In line with this, low-hypodiploidy was far more common than near-haploidy: of the 821 current and former hypodiploid cases (LOH of ≥ 6 autosomes), 96.7% fell within the low-hypodiploid range. Near-haploid cases were identified across 10/33 cancer types but were extremely rare (27/10,332 TCGA cases, 0.26%), reaching a prevalence $\geq 1\%$ only in ACC (6/90), mesothelioma (3/86) and sarcoma (7/248). Most of these cases (24/27) had undergone genome doubling, resulting in a wide range of autosome counts (26–117) at time of sequencing.

The prevalence and distribution of hypodiploid karyotypes varied widely between cancer types (Fig. 1b). Low-hypodiploidy was most common in KICH and adrenocortical carcinomas (ACC), at 80% (52/65 cases) and 54.4% (49/90 cases), respectively. Low-hypodiploidy was rarer in other cancer types but widespread, with 24/33 TCGA cancer types having at least five low-hypodiploid cases and esophageal, lung squamous, ovarian, stomach, lung adenocarcinoma and rectal adenocarcinoma being $\geq 10\%$ low-hypodiploid. Several cancer types other than KICH diverged from the pan-cancer exponential trend, with sarcomas showing a notable surplus of near-haploid cases and ACC presenting a distinctly uniform distribution of mostly genome-doubled low-hypodiploids and near-haploids. Most of the near-haploid sarcomas (4/7) were undifferentiated pleomorphic sarcomas, in which

near-haploidy has been putatively linked to structural variants in the *SMC1A* gene coding for a member of the cohesin complex [67]. The other three cases included two leiomyosarcomas, in which near-haploidisation is thought to occur as an early catastrophic event with non-random retention of chromosomes 5 and 22 [68], and a myxofibrosarcoma. Two blood cancers are represented in the TCGA dataset; no hypodiploids were identified among the acute myeloid leukaemia cases, and hypodiploidy was uncommon in the diffuse B-cell lymphomas (2/47 cases, 4%).

MH score identifies masked hypodiploidy in ALL with high specificity

Genome doubling (WGD) is also common in hypodiploid ALL, resulting in “masked hypodiploids” with hyperdiploid or near-triploid karyotypes resembling high hyperdiploid ALL (HeH-ALL). These cases present a significant diagnostic challenge. Masked hypodiploid acute lymphoblastic leukemia (MH-ALL) retains a very poor prognosis but may be misclassified as HeH-ALL, a subtype with good prognosis and favourable treatment response, if the underlying hypodiploidy remains undetected [11]. Distinguishing these two divergent subsets using only the cytogenetic data available in routine clinical investigations is thus of both biological and clinical significance.

Masked hypodiploids show a characteristic pattern of ‘hyperdiploidy by tetrasomy’, with their ‘hyperdiploid’ karyotype being made up of disomies and tetrasomies (Fig. 2a–b) rather than the disomies and trisomies seen in HeH-ALL [21]. We hypothesized that masked hypodiploid and HeH-ALL could be systematically distinguished by the simple heuristic $\# \text{tetrasomies} - \# \text{trisomies}$, where a masked hypodiploidy (MH) score > 0 indicates MH-ALL and a score ≤ 0 HeH.

To test this heuristic, we used MEDICC2 to construct copy number phylogenies for 150 multi-clone samples from the Mitelman database and identified 126 subclonal masked hypodiploids. For comparison, we obtained allele-specific copy number data for 577 HeH-ALL cases from Woodward et al. (2023) [34] and examined the distribution of MH scores between the two groups. Trisomies were much rarer in masked hypodiploids: the median MH-ALL case had zero trisomies, versus a median of six trisomies in HeH-ALL. There was a strong correlation between the total number of autosomes and the number of trisomies in HeH-ALL ($R = 0.81$) but not MH-ALL ($R = 0.34$) (Fig. 2c). MH scores for the two groups showed little overlap (Fig. 2d), and the heuristic identified masked hypodiploids with a specificity of 99.8% and a sensitivity of 89.7% (Fig. 2e).

This test is simple and applies to both near-haploid and low-hypodiploid ALL. However, it is specific to the

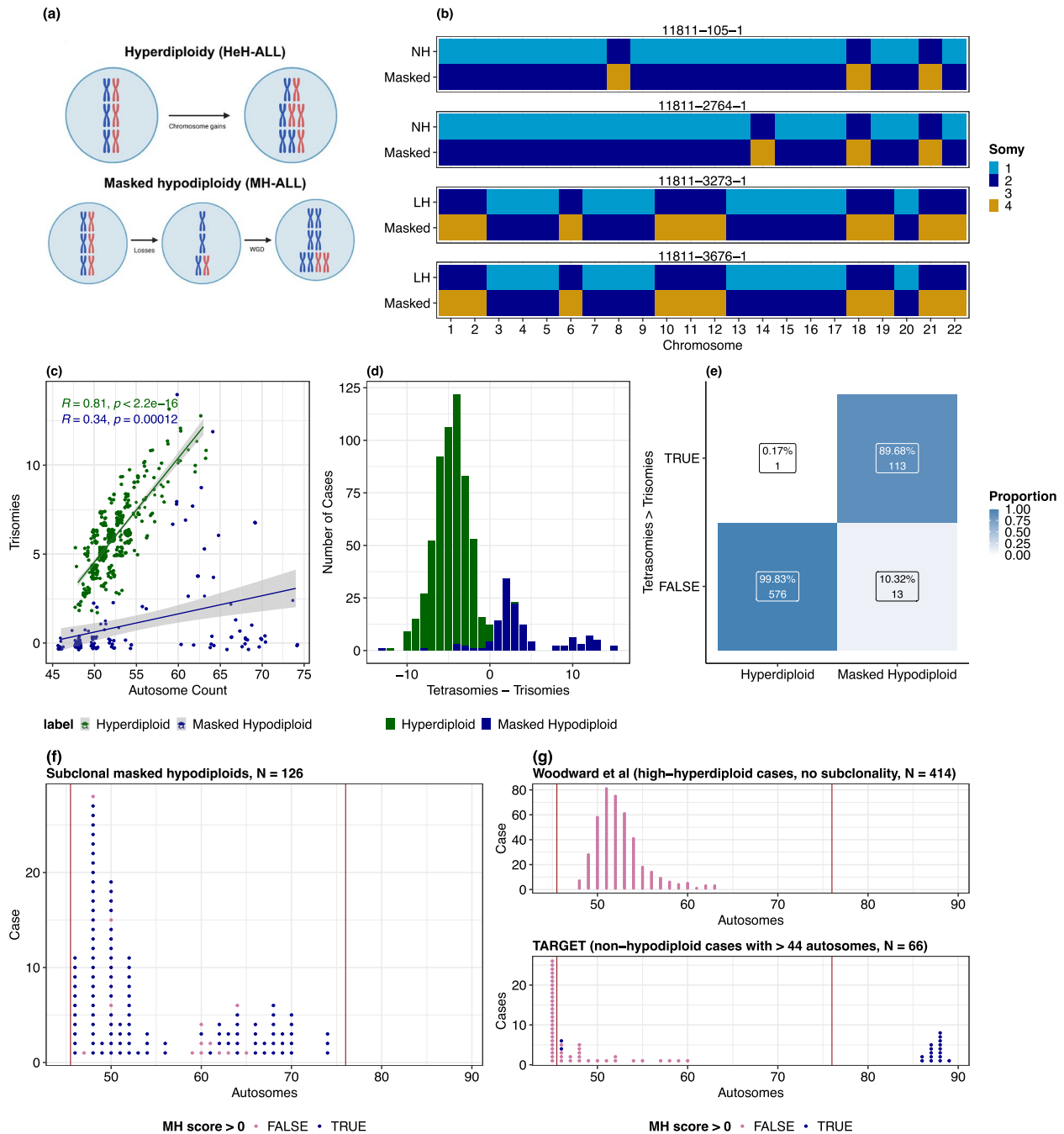


Fig. 2 Identification of masked hypodiploid ALL based on cytogenetic data. **A** Schematic of the origin of hyperdiploid vs masked hypodiploid karyotypes (3 chromosomes visualised). **B** Visualisation of genome doubling leading to subclonal masked hypodiploidy in two low-hypodiploid and two near-haploid ALL cases from the Mitelman database. Patients represented include a two-year-old female, a fifteen-year-old male, a forty two-year old female and a thirteen-year-old female karyotyped by Harrison et al. (2004). NH, near-haploid; LH, low-hypodiploid; HeH, hyperdiploid. **C** Relationship between total autosome count and number of trisomies in hyperdiploid vs masked hypodiploid ALL. Hyperdiploid cases with subclonal whole chromosomes were excluded. Exact p-value for hyperdiploids: $p = 1.506 \times 10^{-97}$. **D** Distribution of MH score (# tetrasomies - # trisomies) in hyperdiploid vs masked hypodiploid ALL. **E** Performance of MH score in distinguishing between high-hyperdiploid and masked hypodiploid ALL. **F** Distribution of autosome counts and MH score classification in subclonal masked hypodiploids from the Mitelman database. **G** Distribution of autosome counts and MH score classification in non-hypodiploid ALL cases: top, high-hyperdiploid cases from Woodward et al. (2023), excluding cases with subclonal variation in chromosome counts; bottom, cases from the TARGET-ALL-P2 project with > 44 autosomes. Vertical lines indicate the range of potential doubled hypodiploid autosome counts

hyperdiploid/near-triploid range and will not detect hypodiploids that have undergone two WGDs, which could be problematic if the likelihood of a first WGD and of a second are not independent. Using allele-specific data for 293 ALL cases from the TARGET-ALL-P2 project, we could not find any examples of this scenario: 0/16 cases in the near-tetraploid range had widespread LOH suggestive of a hypodiploid history (Additional File 1: Fig. S2a). The sensitivity of the MH score heuristic additionally depends on genome stability post-WGD, as gains from disomy and losses from tetrasomy reduce signal. Across the TCGA dataset, which is dominated by chromosomally-unstable tumours, there is a significant overlap between the MH scores of doubled hypodiploids and those of hyperdiploids, reducing discriminative power (Additional File 1: Fig. S2b). In these highly unstable cancer types, allele-specific copy number profiles with LOH information are necessary to reliably detect masked hypodiploidy.

The 13 false negative cases in our analysis (non-hypodiploid and hypodiploid karyotypes in the same sample but MH score ≤ 0) had a median MH score of -3 and 7 trisomies (Additional File 1: Fig. S2c). Four had an MH score of exactly zero. Some of these samples may have been misassigned to the masked hypodiploid test set, i.e. classified as masked hypodiploids when the coexisting hypodiploid and non-hypodiploid populations in the sample were not related by a genome doubling event. However, they may also represent a subtype of MH-ALL that is chromosomally unstable after genome doubling. While low-hypodiploids form a minority of the test masked hypodiploids (45/126, 35.7%), they make up 61.5% (8/13) of the false negative cases, suggesting that low-hypodiploids may be more chromosomally unstable than near-haploids. Most of the false negatives were in cases with chromosome counts in the good-prognosis high-hyperdiploid range, causing sensitivity in this range to fall to 82.5% (Fig. 2f).

MH score detects hidden hypodiploid origins in hyperdiploid ALL cases in the Mitelman database

We used the MH score heuristic to search the Mitelman database for fully-masked hypodiploid ALL cases, i.e. cases where the original hypodiploid clone has been lost and only the genome-doubled clone remains. We identified 13 fully-masked cases in the high-hyperdiploid range (49–65 autosomes); all were clonal except for one case with both a 50-autosome and a 47-autosome subclone. Another 69 hyperdiploid cases possessed a hypodiploid subclone, suggesting that approximately 8.3% of the 989 hyperdiploid ALLs in the Mitelman database may have hypodiploid origins, including 1.3% which are fully masked. This is likely to be an underestimate, given the

high specificity (99.8%) and moderate sensitivity (82.5%) of the MH score in the high-hyperdiploid range.

Broadening our search across the full range of potential doubled hypodiploid autosome counts (46–76), we identified an additional 48 cases (50 clones) with MH score > 0 . Most (46/48) were within the 46–65 autosome range, with only two cases falling in the 66–76 range. This supports the hypothesis that less-extreme hypodiploids rarely undergo WGD or, when they do, the genome-doubled clone seldom becomes fixed within the tumour population. The vast majority of fully-masked cases had autosome counts of 46 (42 clones) or 48 (6 clones) (Additional File 1: Fig. S2d). This differs substantially from the distribution of autosome counts in non-doubled hypodiploid ALL (Fig. 1a) and subclonal genome-doubled hypodiploid ALL (Fig. 2f). In the TARGET-ALL-P2 cohort, only 4/6 (66%) 46-autosome cases were correctly assigned as non-hypodiploid (Fig. 2g). While data in the low-hyperdiploid range is limited, these findings suggest that the specificity of the MH score may decrease at the lower end of the range of potential doubled hypodiploid autosome counts.

Chromosome loss patterns in hypodiploid tumours

Having identified hundreds of current and former hypodiploid tumours across 31 cancer types, we next investigated patterns of chromosome loss in hypodiploids. We analysed all TCGA tumours with a hypodiploid history, but included only currently-hypodiploid ALL cases due to the lack of LOH information in the Mitelman database. Focusing on low-hypodiploid tumours, we found that whole chromosome and chromosome arm loss rates were non-uniform in all 16 cancer types with at least 15 low-hypodiploids (chi-square test of uniformity, $p < 1e-4$ for each cancer type), with the strongest deviations from uniformity seen in ALL, KICH, and ovarian serous cystadenocarcinoma (OV). While we observed some cross-tissue patterns, including frequent retention of chromosomes 7, 1, and 20 and recurrent loss of chromosomes 13 and 17, significant inter-tissue variation remained (Fig. 3a). To investigate the extent to which these non-random chromosome loss and retention patterns in hypodiploid tumours are tissue-specific, we clustered low-hypodiploid samples by their chromosome arm loss profiles (Additional File 1: Fig. S3b). Three cancer types stood out for forming clusters and having higher intra-tissue similarities than their highest inter-tissue similarity: KICH, ALL, and ACC (Fig. 3b). Based on their distinctive tissue-specific loss patterns, we define low-hypodiploids in these cancer types as 'stereotyped hypodiploids'. OV also formed a visible cluster but had slightly more similarity to UCEC than to other OV cases. Other cancer types did not form clear clusters and showed higher median similarity with at least one other

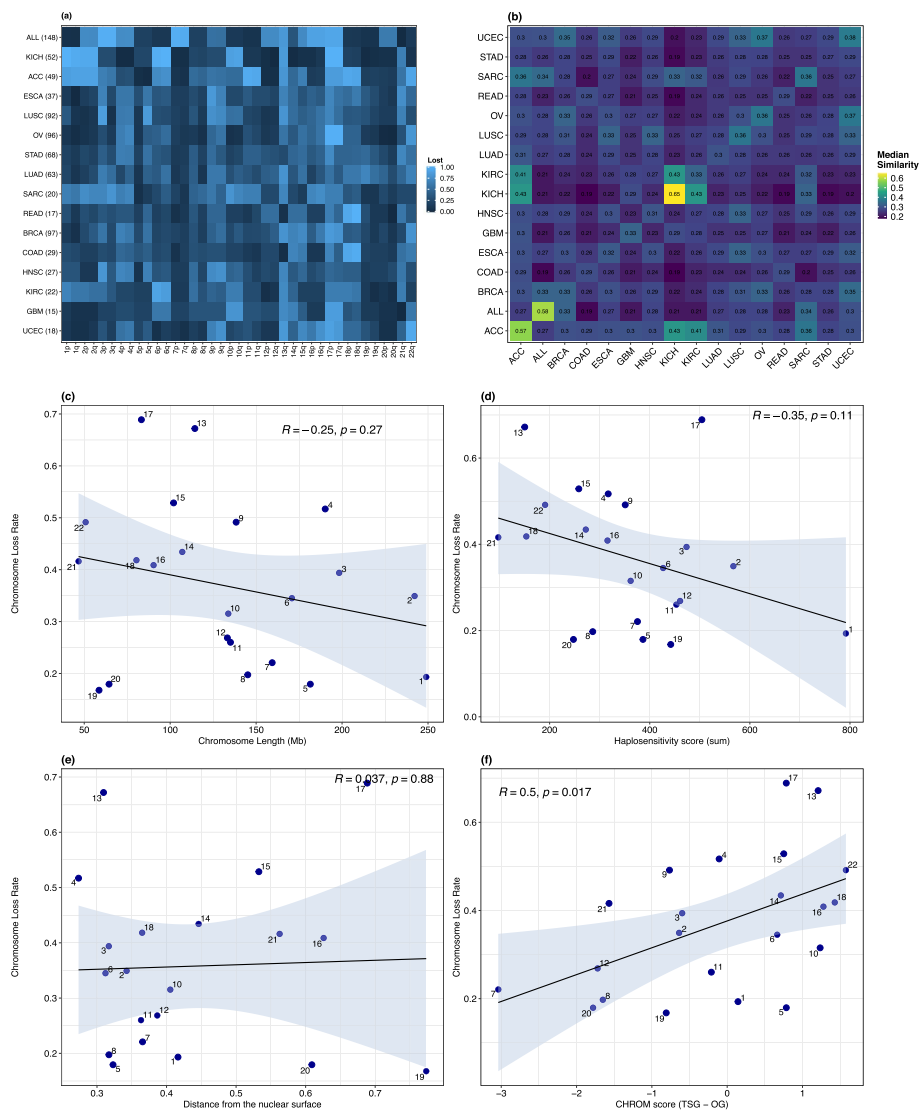


Fig. 3 Patterns of chromosome loss in hypodiploid tumours. **A** Chromosome arm loss rates among low-hypodiploids in cancer types with ≥ 15 low-hypodiploid cases. Number of low-hypodiploid cases per cancer type is shown in brackets. **B** Median pairwise Jaccard similarity between samples of different cancer types. Intra-tissue similarities are shown on the diagonal. Correlation between pan-cancer chromosome loss rates and **C** chromosome length, **D** sum of gene haplosensitivity probabilities derived from Collins et al. (2022), **E** chromosome nuclear position as distance from the nuclear surface, calculated by Girelli et al. (2020) [36] using GPSeq and Hi-C data, **F** weighted CHROM (TSG - OG) score from Davoli et al. [69]. We used LOH data to include both former and current low-hypodiploid TCGA samples. Due to the lack of LOH information in the Mitelman database, we could not precisely identify lost chromosomes in former hypodiploid ALL cases, so we included only currently-hypodiploid ALL cases

tissue than with their own samples. This suggests that, in low-hypodiploids from most cancer types, chromosome loss is largely stochastic and/or driven by factors common across multiple cancer types rather than tissue-specific selection.

Whole-chromosome losses predominated over arm-level events in low-hypodiploids, with strong correlations between loss rates for arms of the same chromosome ($R = 0.78, p = 0.00024$). Inter-arm loss rate correlations in less extreme aneuploid samples were significantly weaker, falling to $R = 0.27$ in tumours with one chromosome loss (Additional File 1: Fig. S3c). Even among

hypodiploid TCGA cases, however, there were outlying chromosomes with diverging loss rates between arms: 5q was lost in 46% of hypodiploid cases while 5p was only lost in 19%, and chromosome 3 showed the opposite pattern, with the p-arm being lost in 53.2% of hypodiploid cases and the q-arm being lost in 31.4%. These patterns suggest selection to retain specific arms despite the overall genomic bias towards chromosomal loss.

It has been reported that chromosome 7 is almost always retained in duplicate even in near-haploid cancers because of selection to maintain heterozygosity of cancer-essential imprinted genes [24]; 7 was indeed the

most-retained chromosome among the low-hypodiploid and near-haploid TCGA tumours, being lost in only 9.74% of cases, but was lost in 95.9% of low-hypodiploid and 100% of near-haploid ALL cases.

Driver gene density is dominant predictor of chromosome loss

To explain non-random chromosome loss patterns in low-hypodiploids, we evaluated chromosome features including length, nuclear location, driver gene density, and dosage sensitivity. Across cancer types, we observed no significant correlation between loss rate and chromosome length (Fig. 3c) or dosage sensitivity scores (Fig. 3d). We next considered the role of chromosome nuclear positioning, using experimentally-derived estimates of chromosome distance from the nuclear surface by Girelli et al. (2020) [36]. Chromosome loss rates in low-hypodiploid tumours were not correlated with distance from the nuclear surface ($r = 0.037$, Fig. 3e).

Of all features tested, driver gene density was the most consistent and significant predictor of chromosome loss. Chromosome-wide TSG-OG density scores derived from Davoli et al. (2013) [69] had a significant positive correlation with chromosome loss rate in hypodiploid tumours ($R = 0.5$, $p = 0.017$, Fig. 3f). In a tissue-specific multivariate regression considering all features, driver gene density was the only significant predictor of chromosome loss rate (Additional File 1: Fig. S3d). In a pan-cancer multivariate analysis, TSG - OG scores were significantly associated with increased loss rate ($p = 0.031$). However, within most individual cancer types, none of the features tested, including driver density, significantly predicted chromosome loss rates.

To determine whether features associated with chromosome loss varied with the degree of aneuploidy, we stratified aneuploid TCGA samples based on the number of chromosomes with LOH (from 1 to 6+, indicating hypodiploid tumours). Haplosensitivity score was associated with a significantly reduced chromosome loss rate in less-extreme aneuploid samples with 1–5 chromosome losses. However, in multivariate analyses, the TSG - OG density score was consistently the only significant predictor of loss rate pan-cancer (Additional File 1: Fig. S3e).

Genome doubling increases with degree of hypodiploidy

Low-hypodiploid tumours are highly aneuploid. Although aneuploidy is strongly associated with chromosomal instability (CIN), the inability to maintain a consistent chromosome complement across cell divisions, some highly aneuploid cancer subtypes such as high-hyperdiploid ALL are chromosomally stable [34]. We thus sought to quantify and characterise the extent of different manifestations of CIN in hypodiploid tumours, including failures of ploidy maintenance (WGD), small-scale

intrachromosomal copy number alterations, structural alterations (chromoanagenesis) and intra-tumour copy number heterogeneity.

Using the MH score heuristic, we first estimated the rate of genome doubling in hypodiploid versus non-hypodiploid ALL in the Mitelman database. WGD was most frequent in near-haploid tumours (67.6%), followed by low-hypodiploids (37%), and almost absent in non-hypodiploid tumours (1.19%) (Fig. 4a). When MH scores for clones with autosome counts outside the well-characterised high-hyperdiploid range were discounted (see Methods), the proportions decreased to 58.8% of near-haploids, 36.1% of low-hypodiploids, and 1.21% of other tumours, but the overall pattern was preserved (Additional File 1: Fig. S4a).

We used MEDICC2 to infer WGD from allele-specific copy number profiles in TCGA and observed a similar trend of increasing WGD rates with increasing degree of hypodiploidy: 88.9% of near-haploids (24/27), 53.4% of low-hypodiploids (424/794), and 32.6% of non-hypodiploid TCGA tumours showed evidence of genome doubling (Fig. 4a). WGD was significantly more prevalent in hypodiploids (LH and NH) than in tumours without a hypodiploid history, even after adjusting for cancer type and tumour purity (logistic regression, $p = 3.53\text{e-}6$). It has been suggested that the link between large-scale loss and WGD is due to selection to buffer LOH of essential regions or normalise dosage [11, 20]. However, an alternative explanation is that both hypodiploidy and genome doubling arise from a shared, generalised chromosomal instability phenotype. The prevalence of low-hypodiploidy and genome doubling is strongly correlated across cancer types ($r = 0.64$, $p = 0.00012$), even after excluding low-hypodiploid cases from the WGD rate calculation (Fig. 4b), a pattern more consistent with a general CIN predisposition varying across cancer types than selection for genome-doubling to restore dosage.

Extensive intrachromosomal instability and heterogeneity in low-hypodiploid tumours

We next considered smaller-scale CIN leading to intrachromosomal copy number alterations (CNAs). Low-hypodiploid tumours have significantly higher numbers of copy number segments across the genome than diploid and less-extreme aneuploid tumours (Fig. 4c; linear regression $p \leq 6.95\text{e-}19$ controlling for cancer type and tumour purity, Additional File 1: Fig. S4b), suggesting a high rate of CNAs beyond the whole-chromosome aneuploidies by which they are defined. An alternative measure of intrachromosomal copy number alterations, the proportion of a chromosome taken up by its longest contiguous copy number segment, shows a similar result, being lower in hypodiploid and polyploid tumours compared to diploids, indicative of more fragmented

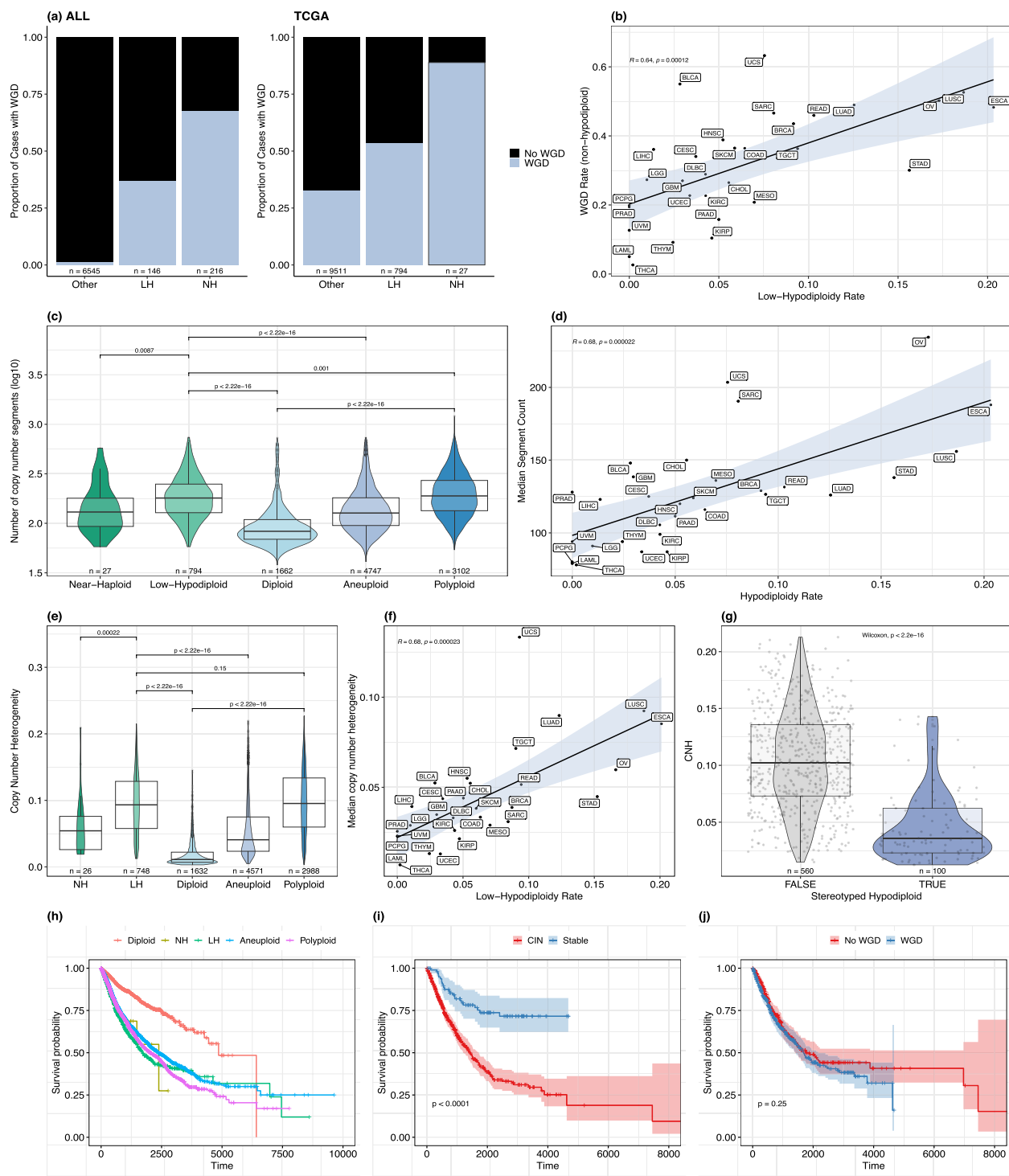


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Hypodiploid tumours are distinguished by chromosomal instability at multiple scales. **A** WGD rate by hypodiploidy level in the Mitelman ALL and TCGA datasets. Multi-clone cases in ALL were assigned their lowest ploidy class. Masked hypodiploidy was detected using the MH score in ALL and using LOH information in TCGA (see Methods). Colours indicate WGD status. **B** Correlation between rate of low-hypodiploidy and WGD rate in non-hypodiploid tumours across TCGA cancer types. **C** Distribution of log₁₀-transformed number of copy number segments by ploidy class. **D** Correlation across TCGA cancer types between rate of low-hypodiploidy and median number of segments per cancer type. Median number of segments was calculated across non-ploidy-altered (diploid and aneuploid) cases. **E** Distribution of copy number heterogeneity (CNH) [6] across ploidy classes. **F** Correlation between rate of low-hypodiploidy and median CNH across TCGA cancer types. Median CNH was calculated across non-ploidy-altered cases. **G** Distribution of CNH in low-hypodiploid cases from cancer types with stereotyped and non-stereotyped chromosome loss patterns (KICH and ACC vs BRCA, LUSC, OV, STAD, LUAD, ESCA, COAD, HNSC, KIRC, SARC, UCEC, READ, GBM); exact p -value = 1.752×10^{-31} . **H** Overall survival by ploidy class. NH, near-haploid; LH, low-hypodiploid. **I** Overall survival in patients with low-hypodiploid tumours from stable (stereotyped) vs CIN-driven cancer types. **J** Overall survival in patients with doubled vs non-doubled low-hypodiploid tumours. P -value calculated using the log-rank test in `ggsurvplot()` is not adjusted for covariates; adjusted p -value is reported in text. Figure 4b–j are based on the TCGA dataset. KICH and ACC were excluded from (b), (d) and (f) due to their extremely high low-hypodiploidy rates. Exact p -values for tests with $p < 2.2 \times 10^{-16}$ are reported in Additional File 2

genomes (Additional File 1: Fig. S4c). These differences remain significant when genome-doubled hypodiploids are excluded from the analysis (Additional File 1: Fig. S4d–e). While hypodiploidy or polyploidy themselves may result in increased intrachromosomal instability, hypodiploidy rate is correlated with median segment count across cancer types ($R = 0.68$, Fig. 4d), further supporting an underlying generalised CIN phenotype.

Hypodiploids are also enriched for structural instability: comparison of our ploidy classes with TCGA chromoanagenesis calls from Rasnic & Linial (2021) [39] indicates that 55.2% of low-hypodiploids show evidence of catastrophic complex chromosomal rearrangements, compared to 8.7% of diploids and 35.8% of less-extreme aneuploids, and second only to polyploid tumours (56.5%) (Additional File 1: Fig. S4f). Chromothripsis is a specific form of chromoanagenesis characterised by clustered chromosome shattering and random restitching. We incorporated chromothripsis calls from 715 TCGA tumours analysed for the Pan-Cancer Analysis of Whole Genomes (PCAWG) project, including 80 low-hypodiploid tumours, and found that low-hypodiploid tumours tended to have a higher number of chromothriptic chromosomes than diploid tumours ($p = 0.0045$), but no significant difference from less-extreme aneuploid tumours ($p = 0.82$) (Additional File 1: Fig. S4g).

To more directly index CIN, we used a copy number heterogeneity (CNH) metric developed by van Dijk et al. (2021) [6], which proxies intra-tumour heterogeneity in copy number profiles based on distance from integer copy number values in bulk tumour samples. Copy number heterogeneity was significantly increased in low-hypodiploid compared to diploid and less-extreme aneuploid tumours, suggesting ongoing karyotypic evolution and diversity in karyotypes between cells of the same tumour (Fig. 4e). This held even when genome-doubled cases were excluded (Additional File 1: Fig. S5a). In line with the strong correlations between hypodiploidy rate and rates of WGD and intrachromosomal CNA, hypodiploidy rate was strongly correlated with median CNH across cancer types (Fig. 4f). In order to account for cancer type confounding, we performed multivariate

linear regression and showed that low-hypodiploids have significantly higher CNH and segment counts than both diploid and less-extreme aneuploid tumours after controlling for cancer type and purity (Additional File 1: Figs. S4b and S5b). Compared to all other tumours, low-hypodiploids show higher CNH ($p = 2.34 \times 10^{-17}$) and segment counts ($p = 0.0214$) even after controlling for their level of aneuploidy as well as their cancer type and purity (Additional File 1: Fig. S5e–f). This reinforces the association between high levels of chromosomal instability and hypodiploidy both between and within cancer types.

Interestingly, near-haploid tumours displayed significantly fewer intrachromosomal copy number alterations ($p < 0.01$, Fig. 4c) and lower intra-tumour copy number heterogeneity ($p < 0.001$, Fig. 4e) than low-hypodiploids, despite having higher rates of WGD. They also showed reduced frequency of chromoanagenesis (33% vs 55%, $p = 0.03$, Fisher's exact test, Additional File 1: Fig. S4f). While near-haploid and low-hypodiploid cases have distinct tissue spectra, near-haploid cases retained significantly lower CNH ($p = 0.011$) and fewer segments (log₁₀-transformed) ($p = 2.23 \times 10^{-3}$) than low-hypodiploids in a linear regression model controlling for cancer type.

Stable hypodiploidy in kidney and adrenocortical tumours diverges from pan-cancer patterns

The degree of chromosomal instability in hypodiploid tumours differs markedly between tissues. We previously identified two cancer types in TCGA (KICH and ACC) with very high rates of hypodiploidy (Fig. 1b) and stereotyped chromosome loss patterns (Fig. 3b). Low-hypodiploids from these cancer types are further distinguished by having significantly lower intra-tumour copy number heterogeneity than low-hypodiploids from cancer types with non-stereotyped chromosome loss patterns (median 0.036 vs 0.103, $p = 1.752 \times 10^{-31}$, Fig. 4g; adjusted for purity, $p = 3.04 \times 10^{-8}$). This could indicate early and/or strongly-selected establishment of a hypodiploid clone with little ongoing copy number evolution or diversification compared to non-stereotyped low-hypodiploids. Moreover, they have significantly fewer intrachromosomal copy number alterations (median 107 vs 194 copy

number segments per genome, $p = 2.924 \times 10^{-26}$, Additional File 1: Fig. S5c; adjusted for purity, $p = 4.6\text{e-}15$), and the longest contiguous copy number segment of their average autosome spans 68% of the chromosome, compared to 48.2% in non-stereotyped hypodiploids.

The MH score heuristic, devised for ALL, is sensitive to karyotypic stability after genome doubling. We used this feature to assess chromosomal instability in hypodiploid tumours that have undergone WGD. We found that the heuristic detects 94.7% of doubled KICH hypodiploids and 87.5% of doubled ACC hypodiploids up to the near-triploid range (≤ 76 autosomes), versus only 19.7% of hypodiploids from non-stereotyped hypodiploid cancer types (Additional File 1: Fig. S5d). These results indicate that hypodiploid ACC and KICH tumours are markedly more stable than hypodiploid cases from non-stereotyped TCGA tumours, even within the subset of cases that have undergone WGD, with stability comparable to hypodiploid ALL.

Hypodiploidy is associated with poor prognosis across cancer types

Although hypodiploidy is an established predictor of poor prognosis in ALL and certain other cancer types including multiple myeloma [70] and Hürthle cell carcinoma [18], its implications for survival in other cancer types are less clear. We assessed overall survival in relation to ploidy status across cancer types and found that overall survival (OS) and progression-free survival (PFS) were highest among patients with diploid tumours, followed by those with less-extreme aneuploid tumours (Fig. 4h, Additional File 1: Fig. S6a). The overall survival of patients with low-hypodiploid tumours tracks closely with that of patients with polyploid tumours, in line with the known relationship between high copy number heterogeneity and poor prognosis [6].

In a multivariate Cox proportional hazards regression controlling for age, sex, race, and cancer type, low-hypodiploid tumours were significantly associated with worse overall survival than both diploid and less-extreme aneuploid tumours (Additional File 1: Fig. S6b). Low-hypodiploid tumours also showed worse progression-free survival than diploid, but not less-extreme aneuploid, tumours (Additional File 1: Fig. S6c). After adjusting for tumour grade at diagnosis, low-hypodiploid tumours were associated with worse OS but not PFS than diploid tumours (Additional File 1: Fig. S6d-e).

Notably, patients with low-hypodiploid tumours from cancer types with stereotyped hypodiploidy (KICH and ACC) had significantly better OS and PFS than those with low-hypodiploid tumours from other cancer types after adjusting for age, sex, race and tumour grade (Fig. 4i, Additional File 1: Fig. S6f-g). This is consistent with the hypothesis that the association between hypodiploidy

and poor prognosis is driven primarily by CIN, rather than a ploidy or genomic imbalance effect.

Additionally, we found no difference in OS or PFS between genome-doubled and non-doubled low-hypodiploid cases (Fig. 4j, Additional File 1: Fig. S6h-i). To test the robustness of these findings, we analysed 1,579 non-TCGA cases from the PCAWG dataset [71], which covers a distinct spectrum of cancer types (Additional File 1: Fig. S6j). These cases recapitulated the survival patterns by ploidy class we observed in the TCGA dataset (Additional File 1: Fig. S6k). We compared patients from the PCAWG dataset with genome-doubled ($N = 49$) versus non-genome-doubled ($N = 46$) low-hypodiploid tumours and found that there was no significant difference in overall survival ($p = 0.32$, Additional File 1: Fig. S6l). When we controlled for patient sex, age and cancer type, genome doubling among low-hypodiploid tumours was associated with better prognosis ($p = 1.77\text{e-}3$). This apparent discrepancy with the TCGA results may reflect differences in cohort composition. In particular, PCAWG low-hypodiploid cases show an overrepresentation of pancreatic endocrine tumours. We speculate that, in certain cancer types, genome doubling in LH tumours may push cells beyond a critical threshold of chromosomal instability, with deleterious consequences [72]. Together, these results support the hypothesis that WGD is enriched in low-hypodiploid tumours not due to a dosage effect increasing the fitness of the tumour but because of their underlying chromosomal instability.

Hypodiploid tumours are enriched for *TP53* mutations and hypoxia

Finally, we sought to shed light on the sources of the genomic instability observed in hypodiploid cancer genomes by analysing the genomic and environmental correlates of hypodiploidy. Overall, low-hypodiploid tumours have higher median mutation counts and ploidy-corrected mutation rates than diploid tumours, but are less likely to be hypermutated (Additional File 1: Fig. S7a-b). Near-haploids did not differ significantly from diploids in mutation count.

We performed multivariate linear regression to detect associations between low-hypodiploidy and mutations in genes from the COSMIC Cancer Gene Census [46] (Fig. 5a). Low-hypodiploid tumours were strongly enriched for *TP53* mutations compared to diploid tumours (log-odds 2.87, $p_{adj} = 6.07\text{e-}31$ controlling for cancer type, mutation burden and purity). Non-synonymous *TP53* mutations occurred in 69.4% of low-hypodiploid cases across the cohort, compared to 54.2% of near-haploids, 52.4% of polyploids, 31.5% of less-extreme aneuploids and only 11.6% of diploid cases (Fig. 5b). Low-hypodiploid tumours were also significantly depleted for mutations in *ARID1A* and *CTCF* ($p_{adj} < 0.05$) compared to

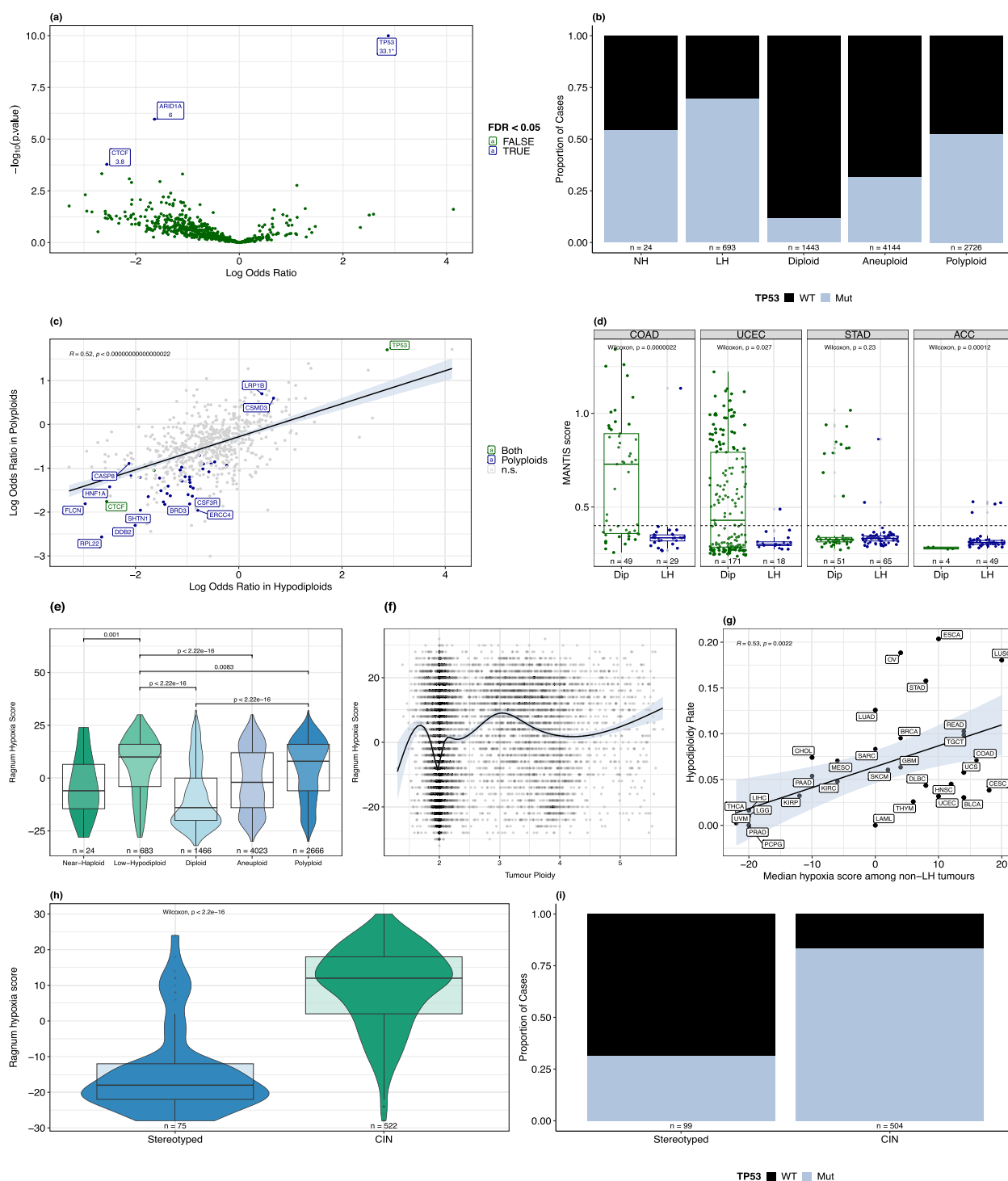


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Hypodiploidy is associated with *TP53* mutations and hypoxia. **A** Genes enriched or depleted for mutations in low-hypodiploid vs diploid tumours, based on logistic regression controlling for total non-synonymous mutation count and cancer type. $-\log_{10}(p\text{-values})$ for genes significant after Benjamini-Hochberg correction are indicated below gene name labels. *TP53*'s $-\log_{10}(p\text{-value})$ is 63.5 but has been capped at 10 for visibility. A group of outliers with estimate ≤ -10 and $p > 0.99$ have been removed. **B** Prevalence of non-synonymous *TP53* mutations by ploidy class. **C** Comparison between regression estimates (log odds ratio) for genes enriched/depleted in low-hypodiploid and polyploid tumours vs diploids; exact $p\text{-value} = 1.14\text{e-}44$. **D** Distribution of microsatellite instability as measured by MANTIS score in colon (COAD), endometrial (UCEC), stomach (STAD) and adrenocortical (ACC) cancers. Dip, diploid; LH, low-hypodiploid. Dashed line indicates threshold for MSI-H status (MANTIS > 0.4). Points are jittered horizontally for visibility. **E** Distribution of Ragnum hypoxia score by ploidy class. **F** Continuous relationship between ploidy and Ragnum hypoxia score. **G** Correlation between median Ragnum hypoxia score in non-hypodiploid tumours and rate of hypodiploidy across TCGA cancer types. **H** Distribution of Ragnum hypoxia score in low-hypodiploid tumours from cancer types with stereotyped (KICH, ACC) vs non-stereotyped ('CIN-driven') hypodiploidy; exact $p\text{-value} = 4.076\text{e-}30$. **I** *TP53* non-synonymous mutation rate in low-hypodiploids from cancer types with stereotyped vs non-stereotyped hypodiploidy. All analyses in this section were performed on the TCGA dataset. Exact $p\text{-values}$ for tests with $p < 2.2\text{e-}16$ are reported in Additional File 2

diploid tumours. The three genes differentially mutated in low-hypodiploid tumours showed similar patterns in polyploid tumours, where they were also significantly enriched or depleted, respectively ($p_{adj} < 0.05$). Moreover, across all COSMIC cancer genes analysed, there was a positive correlation between the log odds ratios of mutations in each gene in low-hypodiploids (vs diploids) and polyploids (vs diploids) overall ($R = 0.52$, Fig. 5c). These results suggest that mutations in these genes affect generalised chromosomal instability or tolerance to aneuploidy rather than WGD specifically. There was no significant difference between low-hypodiploids and less-extreme aneuploid cases in the mutation rate of *ARID1A* or *CTCF*, but low-hypodiploids remained enriched for *TP53* mutations even against other aneuploid cases (estimate 1.83, $p_{adj} = 3.92 \times 10^{-32}$, Additional File 1: Fig. S7c). No genes were significantly differentially mutated in near-haploids compared to low-hypodiploids after multiple testing correction. We additionally used a dataset of pathogenic germline variants identified in TCGA patients by Huang et al. (2018) [28] to search for associations between low-hypodiploidy and germline variation. Out of 18 genes where at least 5 patients carried a germline mutation, no gene was significantly associated with low-hypodiploidy (vs diploidy) after multiple testing correction, although mutations in the Lynch syndrome-associated gene *PMS2* were nominally positively associated with low-hypodiploidy ($p = 0.0195$).

To further characterize mutational patterns in low-hypodiploid tumours, we ran SigProfiler in exome mode to measure COSMIC SBS96 mutational signature activities in TCGA samples based on mutational spectra generated by the PCAWG project [73]. We performed linear regression of the activity of each signature on ploidy class (low-hypodiploids vs diploids), controlling for age, cancer type, tumour purity and total mutation count. Compared to diploid tumours, low-hypodiploid tumours were significantly enriched for the POLD1 defect-associated mutational signature SBS10d ($p_{adj} = 0.0216$) and the unknown signature SBS94 ($p_{adj} = 0.0102$) (Additional File 1: Fig. S7d). There was no significant difference in the prevalence of SBS94 and SBS10d activity between stable

and unstable low-hypodiploid cancer types (Additional File 1: Fig. S7e).

We additionally used TCGA copy number signature attributions from Steele et al. (2022) [74] to compute signature enrichment as the fraction of samples in each ploidy class with non-zero activity for each signature, minus the fraction of all samples (Additional File 1: Fig. S7f). Near-haploids were enriched for some but not all LOH-related signatures (CN14-15) and depleted for the core diploid and tetraploid signatures (CN1 and CN2, respectively). Low-hypodiploids were enriched for CN2 and CN17, which was also common in polyploids and represents complex genomes with a mix of LOH segments in 2–4 copies and heterozygous segments with total copy number between 3–8. This signature appeared in 37.3% of low-hypodiploid samples and 46.7% of polyploids but only 18.4% of less-extreme aneuploid tumours. Non-doubled low-hypodiploids were enriched for CN17 and the LOH-related signature CN9. The enrichment for CN17 was restricted to unstable low-hypodiploids: it was not observed in low-hypodiploid ACC and KICH cases. These cases were more similar to near-haploids, with strong enrichment for the widespread-LOH signature CN14.

Microsatellite instability (MSI) is known to be mutually exclusive with genome doubling in cancer [5, 7, 20]. We extended this observation to low-hypodiploid tumours by analysing microsatellite instability using MANTIS scores from Bonneville et al. (2017) [49]. In all three of the canonical MSI cancer types (colon, endometrial, stomach), low-hypodiploid tumours were less likely to display a high MSI phenotype (MANTIS > 0.4) (Additional File 1: Fig. S7g), and in colon and endometrial cancer, LH tumours had significantly lower MANTIS scores overall (Fig. 5d). This effect was robust to purity (linear regression, $p = 8.52\text{e-}6$ for COAD, $p = 3.6\text{e-}4$ for UCEC). In adrenocortical carcinoma, however, which has both an appreciable rate of microsatellite instability and stereotyped hypodiploidy, hypodiploid tumours have higher MANTIS scores and are more likely to be classified as microsatellite-unstable (Fig. 5d, Additional File 1: Fig. S7g). The same pattern is observed even when

genome-doubled low-hypodiploid tumours are excluded, suggesting that it is not merely a result of the known anti-correlation between microsatellite instability and WGD (Additional File 1: Fig. S7h).

Hypoxia suppresses DNA damage repair responses [75] and induces replication stress and centrosome aberrations [76]. It has been linked to various forms of chromosomal instability in cancer, including polyploidy [77], loss of heterozygosity [74] and chromothripsis [48, 78]. To investigate the relationship between hypoxia and hypodiploidy, we computed hypoxia scores using the Ragnum pimonidazole [47] gene signature across all TCGA samples. Low-hypodiploid tumours were the most hypoxic of any ploidy class, followed by polyploid tumours (Fig. 5e, Additional File 1: Fig. S8a). The increased hypoxia of low-hypodiploid tumours relative to diploid and less-extreme aneuploid tumours was robust to inclusion of cancer type and tumour purity as covariates ($p = 8.94 \times 10^{-15}$ vs less-extreme aneuploids, Additional File 1: Fig. S8b). Low-hypodiploid tumours had significantly higher hypoxia scores than other tumours even after controlling for their level of aneuploidy as well as cancer type and purity (regression $p = 4.88 \times 10^{-9}$, Additional File 1: Fig. S8c).

Interestingly, hypoxia scores varied continuously with ploidy, peaking around ploidy values of 1.5 and 3 (triploidy). Perfectly-diploid tumours were the least hypoxic, with perfectly-tetraploid tumours also having relatively low hypoxia scores (Fig. 5f). Near-haploid tumours were significantly less hypoxic than low-hypodiploid tumours overall, but not after controlling for cancer type (linear regression, $p = 0.287$). To account for the possibility that highly-aneuploid tumours express higher levels of hypoxia-related genes for other reasons, we calculated the median hypoxia score per cancer type and compared it to the rate of hypodiploidy, finding a strong positive correlation ($R = 0.53$, $p = 0.0022$, Fig. 5g). There were significant positive correlations between median hypoxia score and both hypodiploidy and polyploidy across 5 of 8 hypoxia signatures computed by Bhandari et al. (2019) [48] (Additional File 1: Fig. S8c).

We considered the possibility that this relationship is confounded by replication rate, as a high tissue replication rate is associated with polyploidy across cancer types and fast-growing tumours are more likely to become hypoxic. While low-hypodiploidy rate is correlated with median proliferative index across tissues ($R = 0.36$, $p = 0.044$, Additional File 1: Fig. S8d), linear regression of hypoxia score on ploidy class reveals that hypodiploid tumours are still significantly more hypoxic than both diploid and aneuploid tumours after taking proliferative index into account ($p = 1.63 \times 10^{-35}$ and 0.0127 , respectively). After additionally controlling for cancer type, they remain significantly more hypoxic than diploid tumours ($p = 8.93 \times 10^{-11}$) but not less-extreme aneuploids ($p = 0.688$).

Notably, the two TCGA cancer types with stereotyped low-hypodiploid chromosome loss patterns were an exception. Low-hypodiploid cases of kidney chromophobe and adrenocortical cancer had much lower hypoxia than low-hypodiploids from other cancer types (Fig. 5h), and this difference was robust when controlled for tumour purity (linear regression, $p = 1.09 \times 10^{-40}$). These cancer types had the lowest and eighth-lowest median tissue hypoxia levels of all 33 cancer types (Additional File 1: Fig. S8e). Additionally, low-hypodiploid cases of the stereotyped hypodiploid cancers had lower *TP53* mutation rates: *TP53* is non-synonymously mutated in 23.4% of low-hypodiploid ACC and 38.5% of low-hypodiploid KICH, versus a median of 85.7% in low-hypodiploids from other cancer types (Fig. 5i).

Discussion

In recent years, much work has focused on the antecedents, correlates, and outcomes of extreme aneuploidy in cancer, particularly via genome doubling (e.g. [5, 7, 20, 79, 80]). However, outside specific cancer types with stereotyped hypodiploid subtypes (e.g. [8, 13, 18, 81]), the occurrence of extreme aneuploidy and genomic imbalance through chromosome loss has received relatively little attention. Here, we have identified hypodiploidy as a widespread phenomenon across cancer types and systematically characterised patterns of chromosome loss and genomic instability in a pan-cancer framework.

A generalised chromosomal instability phenotype

Beyond their defining characteristic—high levels of whole-chromosome aneuploidy—low-hypodiploid tumours exhibit dramatically increased chromosomal instability at multiple scales, from small-scale intra-chromosomal copy number alterations and chromoanagenesis to whole genome duplication. These tumours also have very high levels of intra-tumour copy number heterogeneity, suggesting ongoing instability rather than stable maintenance of defined hypodiploid karyotypes. While the hypodiploid state itself may promote other forms of chromosomal instability, it has been shown *in vitro* that monosomies alone do not induce CIN [82]. Instead, strong correlations between different forms of CNA observed here and elsewhere [83] suggest that these features are manifestations of an underlying generalised chromosomal instability phenotype.

Like genome-doubled tumours [7], low-hypodiploids are strongly enriched for *TP53* mutations [8] and depleted for microsatellite instability. Low-hypodiploids are also significantly more hypoxic than diploid and less-extreme aneuploid tumours, and rates of both low-hypodiploidy and polyploidy across cancer types correlate with median tissue hypoxia levels. The commonalities between factors associated with low-hypodiploidy and

polyploidy suggest that these factors do not specifically cause genome doubling but instead promote generalised chromosomal instability and/or enable tolerance to extreme aneuploidy.

Intriguingly, near-haploid tumours do not show the same CIN patterns as low-hypodiploids. While our sample size for near-haploids was limited, we observed significantly fewer intrachromosomal copy number events and lower levels of chromoanagenesis and intra-tumour copy number heterogeneity than low-hypodiploid tumours. The prevalence of near-haploidy appears generally unrelated to that of low-hypodiploidy across cancer types. These data support the hypothesis that near-haploidy arises through a distinct mechanism, such as direct near-haploidisation [84], while low-hypodiploids typically result from persistent chromosome missegregation and are thus enriched in cancer types that tend to be chromosomally unstable.

In contrast to the general trend of extreme chromosomal instability among low-hypodiploid tumours, we identified a subset of cancer types with stereotyped low-hypodiploid chromosome loss patterns: ALL, kidney chromophobe, and adrenocortical carcinoma. KICH and ACC were notable for having very high rates of low-hypodiploidy, at 80% and 54.4% compared to 20.3% in the next-highest cancer type, oesophageal carcinoma. Hypodiploidy is an established phenomenon in KICH, with stereotyped loss of chromosomes 1, 2, 6, 10, 13, 17 and, less commonly, 21 [12]. Hypodiploidy was also identified as one of three copy number subtypes in the TCGA-ACC cohort, with frequent genome doubling and preservation of LOH patterns in genome-doubled cases [19].

These stereotyped hypodiploid tumours show significantly reduced intrachromosomal instability, intra-tumour copy number heterogeneity, hypoxia, *TP53* mutation burden and post-WGD missegregation compared to non-stereotyped low-hypodiploids. Hyperdiploid ALL has previously been established as an example of stable aneuploidy [34]; our analysis supports the existence of (relatively) stable extreme aneuploidy after chromosome loss as well. We thus distinguish two classes of hypodiploid tumour: stereotyped hypodiploids, in which hypodiploidy is stable and likely selected for, and sporadic hypodiploids, which represent the extreme end of a spectrum of chromosomal instability. Our analysis suggests that even extreme levels of chromosomal instability rarely result in near-haploid karyotypes or in hypodiploidy rates above ~20%.

The MH score and the diagnostic challenge of masked hypodiploid ALL

Genome doubling presents an important diagnostic challenge in hypodiploid ALL, because it can obscure

poor-prognosis masked hypodiploids with apparent good-prognosis high-hyperdiploid karyotypes, risking treatment failure [11]. Previous research has addressed this problem using loss patterns of specific chromosomes [85] or targeted FISH/flow-cytometry [86, 87], but these approaches either miss near-haploid cases or rely on the persistence of a hypodiploid subclone. Additionally, approaches based on the relative copy numbers of specific chromosomes are likely to suffer from the fact that chromosomes retained in hypodiploids also tend to be gained in hyperdiploids [34].

We thus developed and tested a simple heuristic, the MH score, to distinguish masked hypodiploid from high-hyperdiploid ALL using only cytogenetic information, even in cases of complete masking. The MH score is based on the observation that masked hypodiploids tend to exhibit ‘hyperdiploidy by tetrasomy’, with genomes composed mostly of disomies and tetrasomies, but not trisomies like in hyperdiploid ALL [11, 21]. This pattern is expected following a WGD event on a hypodiploid background. However, chromosomal instability post-doubling can degrade the pattern over time, as missegregation events create additional aneuploidies, including trisomies. Indeed, in more chromosomally unstable tumours, the MH score heuristic loses discriminative power, and masked hypodiploids may be indistinguishable from hyperdiploid/near-triploid cases. It was thus important to test the performance of this heuristic in ALL, and to characterise its sensitivity and specificity in detail across the full range of doubled hypodiploid chromosome counts.

Using a combination of allele-specific copy number data and multi-clone phylogenetic trees, we showed that the MH score is highly specific (>99%) in the clinically relevant high-hyperdiploid autosome count range and sensitive (~90%) across the range of potential doubled hypodiploid autosome counts. Sensitivity falls to 82.5% in the high-hyperdiploid range, and the score heuristic shows low specificity for cases with 46 autosomes, where few missegregation events are required to shift the tetrasomy/trisomy balance. The false negative cases may reflect a chromosomally-unstable subset of ALL or older WGD events. These cases were enriched for low-hypodiploids, supporting the idea that low-hypodiploids are more chromosomally unstable than near-haploids even in ALL. We conclude that the MH score should be a useful tool for analyses of fully-masked hypodiploid ALL using only cytogenetic data. However, we caution that the absence of a ‘hyperdiploidy by tetrasomy’ pattern does not definitively preclude a hypodiploid history, particularly in genomically-unstable cases, and recommend increased adoption of allele-specific genotyping/sequencing in clinical diagnostic contexts where feasible.

The fitness of hypodiploid tumours

Across the TCGA cohort, tumour hypodiploidy is associated with poor prognosis, suggesting increased tumour aggressiveness. This could be explained either by their absolute ploidy and genomic imbalance or by their chromosomal instability. In this work, we have shown that the survival curve for low-hypodiploid tumours overlaps that of polyploid tumours, which have opposite ploidy but similar levels of chromosomal instability. We also demonstrated that the association with poor prognosis is restricted to CIN-driven low-hypodiploids, with stable low-hypodiploid cases having significantly better prognosis. These results dovetail with previous research showing that intra-tumour copy number heterogeneity is a stronger prognostic factor pan-cancer than the mere presence of copy number alterations [6], and that chromosomal instability correlates with disease progression across aneuploid paediatric B-ALL xenografts [88]. The poor prognosis of hypodiploid tumours may therefore be attributable to their underlying genomic instability providing a rich substrate for selection [89].

Extensive chromosome loss and the resulting genomic imbalances are likely to have some effect on the cell, whether by altering signalling networks [90], creating a sensitised background for tumour suppressor gene mutations [18], or inducing metabolic changes [91]. Analysis of chromosome loss patterns in low-hypodiploid tumours reveals no evidence of selection to avoid loss of large or dosage-sensitive chromosomes. The only factor that consistently predicted chromosome loss rate was driver gene density. This finding is consistent with a model in which driver aneuploidies occur early, potentially before the establishment of chromosomal instability, and are followed by unconstrained chromosome losses without regard to the level of genomic imbalance created. Given the inefficiency of negative selection in cancer, particularly in cases with high mutational burdens, it remains unclear to what extent this level of aneuploidy imposes a fitness cost on the tumour [92].

It has been proposed that genome doubling is selected in hypodiploids and in tumours with widespread LOH in order to restore gene dosage and buffer deleterious mutations [11, 20]. However, in low-hypodiploid tumours, we observed no difference in tumour aggressiveness between doubled and non-doubled cases, and suggest that the increased WGD rate in low-hypodiploid tumours may reflect a generalised chromosomal instability phenotype that drives both chromosome loss and genome doubling.

We were not able to assess the relationship between tumour fitness and near-haploidy, but it is plausible that genome doubling is more beneficial to these tumours. Supporting this, we found that genome doubling is strongly enriched in near-haploid tumours despite their significantly lower levels of other forms of chromosomal

instability. However, our MH score analyses indicate that in apparently hyperdiploid ALL cases with evidence of prior hypodiploidy, the genome-doubled or masked hypodiploid clone fully outcompetes the underlying hypodiploid clone in only a minority ($\leq 15\%$) of cases in the Mitelman database. Importantly, near-haploid clones can persist throughout tumour evolution and frequently give rise to highly aggressive tumours and relapses.

Conclusions

Our pan-cancer analysis reveals that most hypodiploid tumours are characterised by extreme chromosomal instability at multiple scales, distinguishes a subset of stable hypodiploids with characteristic chromosome loss patterns, and provides an illustrative example of the extreme freedom from dosage constraint under which tumours evolve.

Abbreviations

ALL	Acute Lymphoblastic Leukaemia
CIN	Chromosomal Instability
CNA	Copy Number Alteration
CPM	Counts Per Million
FDR	False Discovery Rate
HeH	High Hyperdiploid
LOH	Loss of Heterozygosity
LH	Low-Hypodiploid
MAF	Mutation Annotation Format
MH	Masked Hypodiploid
NH	Near-Haploid
OG	Oncogene
OS	Overall Survival
PCAWG	Pan-Cancer Analysis of Whole Genomes
PFS	Progression-Free Survival
TARGET	Therapeutically Applicable Research to Generate Effective Treatments
TCGA	The Cancer Genome Atlas
TSG	Tumour Suppressor Gene
WGD	Whole-Genome Duplication
TCGA	cancer type abbreviations are reported as per the Genomic Data Commons

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-026-01632-y>.

Additional File 1: Supplementary Figures 1-8

Additional File 2: Supplementary Tables 1-10

Acknowledgements

This project is based on data from The Cancer Genome Atlas, Pan-Cancer Analysis of Whole Genomes and TARGET-ALL-P2 projects, as well as the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. The authors thank the patients who participated in these studies and their families. We also thank Tom Kaufmann for helpful discussions.

Authors' contributions

EL conceived of and designed the project, analysed and interpreted data, produced all main and supplemental figures and wrote the manuscript. MNL provided project direction, assisted with data interpretation, edited the manuscript, and provided funding. AMcL supervised the project and provided project direction and funding. All authors read and approved the final manuscript.

Funding

This research was funded by Research Ireland through the Research Ireland Centre for Research Training in Genomics Data Science under grant number 18/CRT/6214 (E.L.), by the European Research Council under grant agreement 771419 (A.McL.), and by the HEA-HCI Pillar 3 initiative (M.N.L. and E.L.).

Data availability

The scripts required to reproduce the analyses in this paper are available at <https://github.com/loughrae/Hypodiploidy> [93]. This paper analyzes existing, publicly available datasets accessible at <https://portal.gdc.cancer.gov> (TCGA and TARGET) and <https://mitelmandatabase.isb-cgc.org> (Mitelman Database of Chromosome Aberrations & Gene Fusions in Cancer).

Declarations

Ethics approval and consent to participate

This paper analyses existing, publicly available datasets.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 September 2025 / Accepted: 23 March 2026

Published online: 22 May 2026

References

- Molina O, Abad MA, Solé F, et al. Aneuploidy in cancer: lessons from acute lymphoblastic leukemia. *Trends Cancer*. 2021;7(1):37–47.
- Taylor AM, Shih J, Ha G, et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell*. 2018;33(4):676–689.e3.
- Klockner TC, Campbell CS. Selection forces underlying aneuploidy patterns in cancer. *Mol Cell Oncol*. 2024;11(1):2369388.
- Nair NU, Schäffer AA, Gertz EM, et al. Chromosome 7 gain compensates for chromosome 10 loss in glioma. *Cancer Res*. 2024;84(20):3464–77.
- Bielski CM, Zehir A, Penson AV, et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet*. 2018;50(8):1189–95.
- van Dijk E, van den Bosch T, Lenos KJ, et al. Chromosomal copy number heterogeneity predicts survival rates across cancers. *Nat Commun*. 2021;12(1):3188.
- Quinton RJ, DiDomizio A, Vittoria MA, et al. Whole-genome doubling confers unique genetic vulnerabilities on tumour cells. *Nature*. 2021;590(7846):492–7.
- Holmfeldt L, Wei L, Diaz-Flores E, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat Genet*. 2013;45(3):242–52.
- McNeer JL, Devidas M, Dai Y, et al. Hematopoietic Stem-Cell transplantation does not improve the poor outcome of children with hypodiploid acute lymphoblastic leukemia: A report from children's oncology group. *J Clin Oncol*. 2019;37(10):780–9.
- Mehta PA, Zhang MJ, Eapen M, et al. Transplantation outcomes for children with hypodiploid acute lymphoblastic leukemia. *Biol Blood Marrow Transplant*. 2015;21(7):1273–7.
- Safari S, Paulsson K. Near-haploid and low-hypodiploid acute lymphoblastic leukemia: two distinct subtypes with consistently poor prognosis. *Blood*. 2017;129(4):420–3.
- Davis CF, Ricketts CJ, Wang M, et al. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell*. 2014;26(3):319–30.
- Olsson L, Paulsson K, Bovée JMG, et al. Clonal evolution through loss of chromosomes and subsequent polyploidization in chondrosarcoma. *PLoS One*. 2011;6(9):e24977.
- Baker TG, Alden J, Dubuc AM, et al. Near haploidization is a genomic hallmark which defines a molecular subgroup of giant cell glioblastoma. *Neuro-Oncol Adv*. 2020. <https://doi.org/10.1093/onoajnl/vdaa155>.
- Hmeljak J, Sanchez-Vega F, Hoadley KA, et al. Integrative molecular characterization of malignant pleural mesothelioma. *Cancer Discov*. 2018;8(12):1548–65.
- Corver WE, Ruano D, Weijers K, et al. Genome haploidisation with chromosome 7 retention in oncocytic follicular thyroid carcinoma. *PLoS One*. 2012;7(6):e38287.
- Bovée JV, van Royen M, Bardoe AF, et al. Near-haploidy and subsequent polyploidization characterize the progression of peripheral chondrosarcoma. *Am J Pathol*. 2000. [https://doi.org/10.1016/s0002-9440\(10\)64796-7](https://doi.org/10.1016/s0002-9440(10)64796-7).
- Gopal RK, Kübler K, Calvo SE, et al. Widespread chromosomal losses and mitochondrial DNA alterations as genetic drivers in hürthle cell carcinoma. *Cancer Cell*. 2018;34(2):242–255.e5.
- Zheng S, Cherniack AD, Dewal N, et al. Comprehensive Pan-Genomic characterization of adrenocortical carcinoma. *Cancer Cell*. 2016. <https://doi.org/10.1016/j.ccell.2016.04.002>.
- López S, Lim EL, Horswell S, et al. Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat Genet*. 2020;52(3):283–93.
- Carroll AJ, Shago M, Mikhail FM, et al. Masked hypodiploidy: hypodiploid acute lymphoblastic leukemia (ALL) mimicking hyperdiploid ALL in children: a report from the children's oncology group. *Cancer Genet*. 2019;238:62–8.
- Nachman JB, Heerema NA, Sather H, et al. Outcome of treatment in children with hypodiploid acute lymphoblastic leukemia. *Blood*. 2007;110(4):1112.
- Casuscelli J, Weinhold N, Gundem G, et al. Genomic landscape and evolution of metastatic chromophobe renal cell carcinoma. *JCI Insight*. 2017;2(12):e92688.
- Boot A, Oosting J, de Miranda NF, et al. Imprinted survival genes preclude loss of heterozygosity of chromosome 7 in cancer cells. *J Pathol*. 2016;240(1):72–83.
- Laurent AP, Kotecha RS, Malinge S. Gain of chromosome 21 in hematological malignancies: lessons from studying leukemia in children with down syndrome. *Leukemia*. 2020;34(8):1984–99.
- Wang J, LaFramboise T. Cytoconverter: a web-based tool to convert karyotypes to genomic coordinates. *BMC Bioinformatics*. 2019;20(1):467.
- Wang J, Zheng J, Lee EE, et al. A cloud-based resource for genome coordinate-based exploration and large-scale analysis of chromosome aberrations and gene fusions in cancer. *Genes Chromosomes Cancer*. 2023;62(8):441–8.
- Huang KL, Mashl RJ, Wu Y, et al. Pathogenic germline variants in 10,389 adult cancers. *Cell*. 2018;173(2):355–370.e14.
- Neph S, Kuehn MS, Reynolds AP, et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics*. 2012;28(14):1919–20.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
- Van Loo P, Nordgard SH, Lingjærde OC, et al. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A*. 2010;107(39):16910–5.
- Colaprico A, Silva TC, Olsen C, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;44(8):e7.
- Kaufmann TL, Petkovic M, Watkins TBK, et al. MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution. *Genome Biol*. 2022;23(1):241.
- Woodward EL, Yang M, Moura-Castro LH, et al. Clonal origin and development of high hyperdiploidy in childhood acute lymphoblastic leukaemia. *Nat Commun*. 2023;14(1):1658.
- Gu Z, Ellis R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847–9.
- Girelli G, Custodio J, Kallas T, et al. GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat Biotechnol*. 2020;38(10):1184–93.
- Barrios D, Prieto C. D3GB: An interactive genome browser for R, python, and WordPress. *J Comput Biol*. 2017;24(5):447–9.
- Collins RL, Glessner JT, Porcu E, et al. A cross-disorder dosage sensitivity map of the human genome. *Cell*. 2022;185(16):3041.
- Rasnic R, Linial M. Chromoanagenesis landscape in 10,000 TCGA patients. *Cancers*. 2021. <https://doi.org/10.3390/cancers13164197>.
- Cortés-Ciriano I, Lee JJK, Xi R, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet*. 2020;52(3):331–41.
- Therneau TM. A Package for Survival Analysis in R. R package version 3.8-6. 2026. <https://CRAN.R-project.org/package=survival>.
- Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.5.2. 2026. <https://rpkgs.datanovia.com/survminer/index.html>.
- Liu J, Lichtenberg T, Hoadley KA, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*. 2018;173(2):400–416.e11.
- Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*. 2015;6(1):8971.

45. Díaz-Gay M, Vangara R, Barnes M, et al. Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment. *Bioinformatics*. 2023;39(12):btad756.
46. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2018;47(Database issue):D941.
47. Ragnun HB, Vlatkovic L, Lie AK, et al. The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer. *Br J Cancer*. 2015. <https://doi.org/10.1038/bjc.2014.604>.
48. Bhandari V, Hoey C, Liu LY, et al. Molecular landmarks of tumor hypoxia across cancer types. *Nat Genet*. 2019;51(2):308–18.
49. Bonneville R, Krook MA, Kautto EA, et al. Landscape of microsatellite instability across 39 cancer types. *JCO Precis Oncol*. 2017;1:1–15.
50. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol*. 2011. <https://doi.org/10.1371/journal.pcbi.1002240>.
51. Barrett T, Dowle M, Srinivasan A, et al. data.table: Extension of 'data.frame'. R package version 1.16.2. 2024. <https://github.com/rdatatable/data.table>.
52. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *Journal of Open Source Software*. 2019;4(43):1686. <https://doi.org/10.21105/joss.01686>.
53. Firke S. janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.2.0. 2024. <https://github.com/sfirke/janitor>.
54. Ramos M, Schiffer L, Waldron L. TCGAutils: TCGA utility functions for data management. R package version 1.31.5. 2026. <https://bioconductor.org/packages/TCGAutils>.
55. Schaubberger P, Walker A. openxlsx: Read, Write and Edit xlsx Files. R package version 4.2.7. 2025. <https://ycphs.github.io/openxlsx/index.html>.
56. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0. 2026. <https://rpkgs.datanovia.com/ggpubr/>.
57. Slowikowski K. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.6. 2026. <https://github.com/slowkow/gg-repel>.
58. Pedersen TL. patchwork: The Composer of Plots. R package version 1.3.0. 2025. <https://patchwork.data-imaginat.com>.
59. Yu G. ggplotify: Convert Plot to 'grob' or 'ggplot' Object. R package version 0.1.2. 2023. <https://CRAN.R-project.org/package=ggplotify>.
60. Neuwirth E. RColorBrewer: ColorBrewer Palettes. R package version 1.1–3. 2022. <https://CRAN.R-project.org/package=RColorBrewer>.
61. Robinson D, Hayes A, Couch S, et al. broom: Convert Statistical Objects into Tidy Tibbles. R package version 1.0.7. 2025. <https://broom.tidymodels.org/>.
62. Signorell A. DescTools: Tools for Descriptive Statistics. R package version 0.99.58. 2025. <https://andrisignorell.github.io/DescTools/>.
63. Johnston B. figpatch: Easily Arrange External Figures with Patchwork Alongside 'ggplot2' Figures. R package version 0.3.0. 2025. <https://github.com/BradYJohnston/figpatch>.
64. Barnett DJ, Arts IC, Penders J. microViz: an R package for microbiome data visualization and statistics. *Journal of Open Source Software*. 2021;6(63):3201. <https://doi.org/10.21105/joss.03201>.
65. Alaghebandan R, Przybycin CG, Verkarre V, et al. Chromophobe renal cell carcinoma: Novel molecular insights and clinicopathologic updates. *Asian J Urol*. 2021;9(1):1.
66. Di Mauro I, Ambrosetti D, Vignot L, et al. Detection of tetraploidization in chromophobe renal cell carcinoma: insights and pitfalls. *Genes Chromosomes Cancer*. 2020;59(12):675–87.
67. Ibstedt S, Piccinelli P, Sydow S, et al. Structural variants in the SMC1A gene associated with near-haploidy in undifferentiated pleomorphic sarcomas. *Genes, Chromosomes and Cancer*. 2024. <https://doi.org/10.1002/gcc.23255>.
68. Arbajian E, Köster J, Vult von Steyrn F, et al. Inflammatory leiomyosarcoma is a distinct tumor characterized by near-haploidization, few somatic mutations, and a primitive myogenic gene expression signature. *Mod Pathol*. 2018;31(1):93–100. <https://doi.org/10.1038/modpathol.2017.113>.
69. Davoli T, Xu AW, Mengwasser KE, et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell*. 2013;155(4):948–62.
70. Smadja NV, Bastard C, Brigaudeau C, et al. Hypodiploidy is a major prognostic factor in multiple myeloma. *Blood*. 2001. <https://doi.org/10.1182/blood.v98.7.2229>.
71. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82–93.
72. Hosea R, Hillary S, Naqvi S, et al. The two sides of chromosomal instability: drivers and brakes in incancer. *Signal Transduct Target Ther*. 2024;9(1):75.
73. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94–101.
74. Steele CD, Abbasi A, Islam SMA, et al. Signatures of copy number alterations in human cancer. *Nature*. 2022. <https://doi.org/10.1038/s41586-022-04738-6>.
75. Suvac A, Ashton J, Bristow RG. Tumour hypoxia in driving genomic instability and tumour evolution. *Nat Rev Cancer*. 2025;25(3):167–88.
76. Mittal K, Kaur J, Sharma S, et al. Hypoxia drives centrosome amplification in cancer cells via HIF1 α -dependent induction of Polo-Like kinase 4. *Mol Cancer Res*. 2022;20(4):596–606.
77. Lopez-Sánchez LM, Jimenez C, Valverde A, et al. CoCl₂, a mimic of hypoxia, induces formation of polyploid giant cells with stem characteristics in colon cancer. *PLoS One*. 2014;9(6):e99143.
78. Bhandari V, Li CH, Bristow RG, et al. Divergent mutational processes distinguish hypoxic and normoxic tumours. *Nat Commun*. 2020;11(1):1–10.
79. Baker TM, Lai S, Lynch AR, et al. The history of chromosomal instability in genome-doubled tumors. *Cancer Discov*. 2024;14(10):1810–22.
80. Burdett NL, Willis MO, Pandey A, et al. Timing of whole genome duplication is associated with tumor-specific MHC-II depletion in serous ovarian cancer. *Nat Commun*. 2024;15(1):1–13.
81. Michal M, Kravtsov O, Ross JS, et al. Clear cell mesotheliomas with inactivating VHL mutations and near-haploid genomic features. *Genes, Chromosomes and Cancer*. 2023. <https://doi.org/10.1002/gcc.23119>.
82. Hintzen DC, Soto M, Schubert M, et al. The impact of monosomies, trisomies and segmental aneuploidies on chromosomal stability. *PLoS One*. 2022;17(7):e0268579.
83. Xian S, Dosset M, Almanza G, et al. The unfolded protein response links tumor aneuploidy to local immune dysregulation. *EMBO Rep*. 2021;22(12):e52509. <https://doi.org/10.15252/embr.202152509>.
84. Walen KH. Haploidization of human diploid metaphase cells: is this genome reductive mechanism operational in near-haploid leukemia? *J Cancer Ther*. 2013;5(1):101–14.
85. Creasey T, Enshaei A, Nebral K, et al. Single nucleotide polymorphism array-based signature of low hypodiploidy in acute lymphoblastic leukemia. *Genes Chromosomes Cancer*. 2021. <https://doi.org/10.1002/gcc.22956>.
86. Chaturvedi A, Shetty D, Ghogale SG, et al. Detecting hypodiploidy with endoreduplication and masked hypodiploidy in b-cell acute lymphoblastic leukemia using multicolor flow cytometry. *Cytometry B Clin Cytom*. 2022. <https://doi.org/10.1002/cyto.b.22063>.
87. Gupta T, Arun SR, Babu GA, et al. A systematic cytogenetic strategy to identify masked hypodiploidy in precursor B acute lymphoblastic leukemia in low resource settings. *Indian J Hematol Blood Transfus*. 2021. <https://doi.org/10.1007/s12288-021-01409-w>.
88. Molina O, Ortega-Sabater C, Thampi N, et al. Chromosomal instability in aneuploid acute lymphoblastic leukemia associates with disease progression. *EMBO Mol Med*. 2024;16(1):64–92.
89. Bakhoun SF, Cantley LC. The multifaceted role of chromosomal instability in cancer and its microenvironment. *Cell*. 2018;174(6):1347–60.
90. Zhakula-Kostadinova N, Taylor AM. Patterns of aneuploidy and signaling consequences in cancer. *Cancer Res*. 2024. <https://doi.org/10.1158/0008-5472.can-24-0169>.
91. Magesh RY, Kaur AN, Keller FN, et al. Aneuploidy generates enhanced nucleotide dependency and sensitivity to metabolic perturbation. *Genes Dev*. 2025;39(11–12):770–86.
92. Tilik S, Tkachenko S, Curtis C, et al. Most cancers carry a substantial deleterious load due to Hill-Robertson interference. *eLife Sciences Publications Limited*. 2022.
93. Loughran E. Code for: Pervasive chromosomal instability drives the karyotypic evolution of hypodiploid tumours. Github. 2026. <https://github.com/lo-ughrae/Hypodiploidy>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.