

Inaugural dissertation
for
obtaining the doctoral degree of
the
Combined Faculty of Mathematics, Engineering and Natural
Sciences of the
Ruprecht - Karls - University
Heidelberg

Presented by
M.Sc. George Philippos
Born in: Aleppo, Syria
Oral examination: 28.05.2025

**Investigating the mechanistic basis
of chromothripsis initiation and the clonal evolution
of chromothriptic cells**

Referees:

Prof. Dr. Oliver Stegle

PD Dr. Aurélie Ernst

Declaration

The work presented in this dissertation was performed from February 2021 to March 2025 under the supervision of PD Dr. Aurélie Ernst at the German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ) in Heidelberg, Germany.

I hereby declare that the thesis I submitted is my own work and has been written independently by me. In the process, I have not used any sources or materials other than those explicitly stated. I confirm that all data in this dissertation are original unless stated and referenced otherwise. I conducted the experiments myself and the specific contributions of others are clearly acknowledged in the text, figure legends and acknowledgements section. Furthermore, I hereby declare that I have not applied to be examined at any other institution, nor have I submitted this thesis, in whole or in part, as an examination paper or as a dissertation to any other faculty or institution.

Abstract

Chromothripsis is a form of genomic instability characterized by massive chromosomal shattering, resulting in complex rearrangements. This phenomenon affects 30-50% of all cancers and is strongly linked to poor prognosis. *TP53* mutations are tightly associated with chromothripsis.

Although several mechanisms have been proposed, the specific triggers, molecular processes driving chromothripsis and the precise sequence of events are still unclear. In particular, how *TP53* loss exactly drives replication stress and genomic instability, ultimately leading to such catastrophic events, is poorly understood. Furthermore, the timing of telomere stabilization, which is essential for chromothriptic cells to survive and evade apoptosis, remains unresolved. Moreover, our current knowledge of chromothripsis initiation is largely based on artificially induced *in vitro* systems or retrospective genomic analyses of fully developed tumours.

To address these challenges, I used skin-derived fibroblasts from Li-Fraumeni Syndrome (LFS) patients, which undergo spontaneous chromothripsis. In these fibroblasts, bulk whole genome sequencing (WGS) data showed chromothripsis at late passages (*TP53*^{-/-}), whereas the matched early-passage (*TP53*^{+/-}) cells did not show clonal chromothripsis. Through a combination of phenotypic assays and cutting-edge multiomics approaches, I performed a longitudinal study to investigate two major questions: i) what leads to chromothripsis, and ii) what gives chromothriptic cells a selective advantage to become dominant in a cell population.

The longitudinal profiling revealed that as early-passage cells advanced towards the growth crisis, they lost the wildtype *TP53* allele, thus triggering global hypertranscription and increased nucleotide consumption. Simultaneously, dysfunctional p53 impaired the pentose phosphate pathway, leading to reduced nucleotide synthesis, as evidenced by transcriptomic and proteomic analyses. This imbalance resulted in replication stress induced by a competition between transcription and replication for the limited nucleotide pool. As a consequence of replication stress and checkpoint dysregulation due to p53 functional loss, LFS cells exhibited increased telomere attrition, breakage-fusion-bridge (BFB) cycles, chromatin bridges, multipolar spindle formation, chromosome missegregation, and micronuclei formation. This cascade of events ultimately led to chromothripsis and massive complex rearrangements. In addition, Single-cell DNA analysis revealed a high heterogeneity across cells, along with frequent copy number losses, rare events and diverse chromothriptic events already at early passages. Conversely, post-crisis chromothriptic cells exhibited clonal chromothripsis and complex rearrangements.

Notably, while most cells underwent apoptosis or negative selection, I found that a small subset of cells survived due to selective advantages and the activation of key adaptation mechanisms.

These included telomere stabilization via alternative lengthening of telomeres (ALT) or *TERT* activation, extrachromosomal circular DNA (ecDNA) formation, gene fusions, oncogene activation, and loss of tumour suppressor genes. The activation of such mechanisms conferred the cells survival advantages, driving rapid expansion and clonal evolution. Thus, I captured the sequence of events, from the non-neoplastic cell population to the appearance of dominant clones carrying chromothriptic chromosomes.

Taken together, this study unravels the complex interplay between p53 dysfunction, nucleotide metabolism, and genomic instability in the initial phases of chromothripsis, thus providing crucial insights into the first steps of chromothripsis-mediated cancer development. Consequently, these findings may guide future strategies for cancer prevention and therapeutic interventions.

Zusammenfassung

Chromothripsis ist eine Form der genomischen Instabilität, die durch massive chromosomale Umstrukturierungen gekennzeichnet ist und die zu komplexen Umlagerungen führt. Dieses Phänomen betrifft 30-50 % aller Krebsarten und ist stark mit einer schlechten Prognose verbunden. *TP53*-Mutationen stehen in engem Zusammenhang mit Chromothripsis.

Obwohl mehrere Mechanismen vorgeschlagen wurden, sind die spezifischen Auslöser und molekularen Prozesse, die Chromothripsis antreiben, sowie die genaue Abfolge der Ereignisse noch immer unzureichend verstanden. Insbesondere ist unklar, wie der Verlust von *TP53* den Replikationsstress und die genomische Instabilität antreibt, was letztlich zu solch katastrophalen Ereignissen führt. Zusätzlich bleibt der Zeitpunkt der Telomerstabilisierung, deren Entkommen vor der Apoptose essenziell ist, ungelöst. Darüber hinaus basiert unser aktuelles Wissen über die Auslösung der Chromothripsis größtenteils auf künstlich induzierten *In-vitro*-Systemen oder retrospektiven genomischen Analysen vollständig entwickelter Tumoren.

Um diese Herausforderungen zu bewältigen, verwendete ich Fibroblasten aus der Haut von Patienten mit Li-Fraumeni-Syndrom (LFS), bei welchen es zu spontaner Chromothripsis kommt. Bei diesen Fibroblasten zeigten die Daten der Gesamtgenomsequenzierung (WGS) Chromothripsis bei späten Passagen (*TP53*^{-/-}), während die entsprechenden Zellen bei frühen Passagen (*TP53*^{+/-}) keine klonale Chromothripsis aufwiesen. Durch eine Kombination aus phänotypischen Tests und hochmodernen Multiomics-Analysen führte ich eine longitudinale Studie durch, um zwei Hauptfragen zu untersuchen: i) Was führt zu Chromothripsis? und ii) Welche selektiven Vorteile ermöglichen chromothriptischen Zellen, in einer Zellpopulation dominant zu werden?

Die longitudinale Profilierung ergab, dass Zellen in der frühen Passage beim Fortschreiten in die Wachstumskrise das wildtypische *TP53*-Allel verloren und dadurch eine globale Hypertranskription und einen erhöhten Nukleotidverbrauch auslösten. Gleichzeitig beeinträchtigte das funktionsgestörte p53 den Pentosephosphatweg, was zu einer reduzierten Nukleotidsynthese führte, die durch transkriptomische und proteomische Analysen nachgewiesen wurde. Dieses Ungleichgewicht führte zu Replikationsstress, der durch den Wettbewerb zwischen Transkription und Replikation um den begrenzten Nukleotidpool induziert wurde. Als Folge des Replikationsstresses und der Checkpoint-Dysregulation aufgrund des Verlusts der p53-Funktion wiesen LFS-Zellen eine erhöhte Telomerverkürzung, Breakage-Fusion-Bridge (BFB)-Zyklen, Chromatinbrücken, multipolare Spindelbildung, Chromosomenfehlsegregation und Mikronuklei-Bildung auf. Diese Kaskade von Ereignissen führten schließlich zu Chromothripsis und massiven komplexen Umlagerungen. Darüber hinaus zeigte die Einzelzell-DNA-Analyse eine hohe Heterogenität zwischen den Zellen,

zusammen mit häufigen Verlusten der Kopienzahl, seltenen Ereignissen und vielfältigen chromothriptischen Ereignissen bereits in frühen Passagen. Im Gegensatz dazu wiesen post-Krise-chromothriptische Zellen klonale Chromothripsis und komplexe Umlagerungen auf.

Bemerkenswerterweise, Während die Mehrheit der Zellen der Apoptose oder negativen Selektion unterlag, stellte ich fest, dass eine kleine Untergruppe aufgrund selektiver Vorteile und der Aktivierung zentraler Anpassungsmechanismen überlebte. Diese umfassten die Telomerstabilisierung durch alternative Telomerverlängerung (ALT) oder *TERT*-Aktivierung, Bildung extrachromosomaler zirkulärer DNA (ecDNA), Genfusionen, Onkogenaktivierung und Verlust von Tumorsuppressorgenen. Die Aktivierung dieser Mechanismen verschaffte den Zellen Überlebensvorteile, was zu einer schnellen Expansion und klonale Evolution führte. Somit erfasste ich die Abfolge von Ereignissen, von den nicht-neoplastischen Zellpopulationen bis zum Auftreten dominanter Klone mit chromothriptischen Chromosomen.

Zusammenfassend entschlüsselt diese Studie das komplexe Zusammenspiel zwischen p53-Dysfunktion, Nukleotid-Stoffwechsel und genomischer Instabilität in den frühesten Phasen der Chromothripsis und liefert somit entscheidende Erkenntnisse über die ersten Schritte der chromothripsis-vermittelten Krebsentstehung. Folglich könnten diese Ergebnisse zukünftige Strategien zur Krebsprävention und therapeutischen Interventionen leiten.

Table of Contents

Declaration	i
Abstract	iii
Zusammenfassung	v
List of figures	xi
List of tables	xiii
Abbreviations	xiv
1. Introduction	1
1.1. Genomic instability and cancer development	1
1.2. The definition of chromothripsis	2
1.3. Chromothripsis prevalence across human cancers	3
1.4. Characteristics of chromothripsis	5
1.5. Key criteria for identifying chromothripsis	6
1.6. Methods to detect chromothripsis and CNAs	7
1.7. Mechanisms underlying chromothripsis	9
1.7.1. Telomere dysfunction and BFB cycles	9
1.7.2. Micronuclei formation	10
1.7.3. Aberrant mitosis	12
1.7.4. Replication stress	14
1.7.5. Additional proposed mechanisms of chromothripsis	14
1.8. Consequences of chromothripsis and selective advantages	15
1.9. Clinical implications of chromothripsis	16
1.10. Current models to study chromothripsis	17
1.11. Li-Fraumeni syndrome as a paradigm to study chromothripsis	18
2. Aim of the study	20
3. Results	21
3.1. LFS <i>in vitro</i> model	21
3.2. Morphology and growth characteristics of LFS fibroblasts at different time points	22
3.3. Chromatin bridges and micronuclei as phenotypic characteristics of genomic instability	24
3.4. Characterization of mitotic defects in LFS cells	26
3.5. Comprehensive DNA profiling and SV analysis	27
3.5.1. Comparative analysis of SV profiles using Strand-seq	27
3.5.2. Dissecting evolutionary dynamics using SV-based clustering	28
3.5.3. Unravelling genomic instability and diversity across cell passages	30
3.5.4. Mapping clonal dynamics and their evolutionary trajectories using SVs	31

3.5.5. Chromothripsis and complex rearrangements at the single-cell level	34
3.5.6. Characterizing ploidy states in Strand-seq and metaphase spreads	35
3.5.7. High-throughput sequencing reveals key CNAs	37
3.5.8. Bulk WGS analysis identifies chromothriptic chromosomes and key drivers in LFS	39
3.6. Telomere stabilization in LFS fibroblasts	43
3.7. Inducing cellular immortalization via <i>hTERT</i> transduction	45
3.7.1. Accelerated growth and decreased stress morphology in <i>hTERT</i> ⁺ cells, with no signs of crisis	46
3.7.2. Significant reduction in chromatin bridges and micronuclei in LFS cells following <i>hTERT</i> transduction	47
3.7.3. Single-cell sequencing analysis reveals reduced genomic instability in <i>hTERT</i> ⁺ immortalized cells	48
3.8. Drivers of dominant clones	49
3.8.1. ecDNAs as a defining feature of winning clones	50
3.8.2. Gene fusion detection through bulk genomic and transcriptomic analysis	51
3.9. Temporal profiling using mass spectrometry-based proteomics	52
3.9.1. Gene set enrichment and pathway analysis of differentially expressed proteins ..	53
3.9.1.1. Identification of key biological processes using GSEA GO	53
3.9.1.2. Complementary pathway insights with GSEA Reactome and GSEA KEGG ..	57
3.9.1.3. Analysis of protein levels in key GSEA GO processes and pathways	59
3.9.2. Identifying transcriptional regulators through ChEA3 analysis	61
3.10. Transcriptome analysis of LFS fibroblasts	61
3.10.1. Investigation of critical GO terms from bulk RNA-seq associated with chromothripsis and clonal expansion	63
3.10.2. Expanded GSEA pathway analysis using Reactome and KEGG databases	65
3.10.3. Exploring DEG patterns across key GSEA GO processes and pathways	67
3.11. Role of hypertranscription and nucleotide depletion in replication stress	69
3.12. Analysis of hypertranscription in cancer using TCGA data	70
4. Discussion	73
4.1. LFS skin-derived fibroblasts as a spontaneous model to study chromothripsis	73
4.2. LFS cells exhibit altered morphology as they progress to crisis and post-crisis passages	74
4.3. LFS cells show early chromosomal bridges, multipolar spindles, and micronuclei, which increase during the crisis	76
4.4. High SV diversity already at early passages and later selection of dominant clones ..	78
4.5. Single-cell Strand-seq analysis reveals chromothripsis, complex rearrangements and aneuploidy already in early-passage cells	79

4.6. HIPSD-seq analysis of CNAs uncovers key factors driving chromothripsis, cellular survival and progression	80
4.7. Bulk WGS identifies recurrent chromosomes affected by chromothripsis and genetic drivers in LFS fibroblasts	81
4.8. Telomere stabilization occurs close to the crisis and is essential for cell survival.....	81
4.9. Telomerase activation prevents crisis and reduces chromosomal abnormalities	82
4.10. ecDNAs and gene fusions confer chromothriptic cells selective advantages.....	83
4.11. Proteomics and transcriptomics data reveal that hypertranscription and metabolic dysregulation result in impaired nucleotide pools and subsequent replication stress.....	84
4.12. Limitations and future perspectives	87
5. Conclusion	89
6. Materials and methods	90
6.1. Materials	90
6.1.1. Cell lines.....	90
6.1.2. Oligos used for HIPSD-seq.....	90
6.1.3. R packages	91
6.1.4. Python packages	91
6.1.5. Softwares	92
6.1.6. Databases and online tools	92
6.2. Methods.....	92
6.2.1. Cell culture	93
6.2.2. Cell morphology imaging and analysis.....	94
6.2.3. Growth curves for LFS fibroblasts.....	94
6.2.4. Quantification of micronuclei and chromatin bridges	94
6.2.5. Immunofluorescence analysis of pH3 and Acetyl-alpha-Tubulin and quantification of mitotic defects	94
6.2.6. <i>TERT</i> -FISH.....	95
6.2.7. ALT FISH	96
6.2.8. Quantification of ecDNAs and ploidy assessment.....	97
6.2.9. Retroviral transduction protocol of <i>hTERT</i>	98
6.2.10. Cumulative PDLs of transduced cells	99
6.2.11. Bulk RNA-sequencing.....	99
6.2.11.1. RNA extraction and sequencing	99
6.2.11.2. bulk RNA-seq data processing and alignment.....	100
6.2.11.3. Differential gene expression and GSEA analyses.....	100
6.2.12. Proteomics analysis of LFS fibroblasts	101
6.2.12.1. Sample preparation for mass spectrometry analysis	101

6.2.12.2. Mass spectrometry data acquisition	102
6.2.12.3. Mass spectrometry data processing analysis and visualization	102
6.2.12.4. Downstream proteomic GSEA and Cytoscape analyses	103
6.2.12.5. Transcription factor enrichment analysis	103
6.2.13. Bulk whole-genome sequencing (WGS)	103
6.2.13.1. DNA extraction and sequencing	103
6.2.13.2. Analysis of bulk whole-genome sequencing data	104
6.2.13.3. OncoAnalyser	104
6.2.13.4. AmpliconSuite	104
6.2.13.5. Chromothripsis detection.....	105
6.2.14. Single-cell DNA sequencing by HIPSD-seq and HIPSD&R-seq	105
6.2.14.1. Tn5 loading	105
6.2.14.2. HIPSD-seq and HIPSD&R-seq sample preparation	106
6.2.14.3. Nuclei extraction and nucleosome depletion	106
6.2.14.4. HIPSD-seq library preparation.....	106
6.2.14.5. HIPSD&R-seq library preparation.....	107
6.2.14.6. HIPSD-seq and HIPSD&R-seq computational analysis	107
6.2.15. Strand-seq.....	108
6.2.15.1. Strand-seq library preparation and sequencing	108
6.2.15.2. Strand-seq data processing	110
6.2.15.3. Clustering of single cells based on SVs.....	110
6.2.15.4. Fold-back inversions calling (BFB patterns) in Strand-seq data.....	111
6.2.16. Primary Template-directed Amplification (PTA)	111
6.2.16.1. FACS sorting.....	111
6.2.16.2. Whole genome amplification using PTA protocol	112
6.2.16.3. PTA library preparation and sequencing.....	113
6.2.16.4. PTA bioinformatics analysis.....	114
6.2.17. Analysis of hypertranscription	114
6.2.18. Statistics and reproducibility	115
6.2.19. Thesis writing.....	115
Supplementary figures.....	116
Publications	132
Acknowledgements	133
References.....	135

List of figures

Figure 1. Chromothripsis and the evolutionary consequences of genomic chaos	2
Figure 2. Chromothripsis frequency across human cancers	4
Figure 3. scDNA sequencing by Strand-seq and HIPSD&R-seq to detect and study genomic instability and chromothripsis.....	8
Figure 4. Mechanisms underlying BFB cycles and their role in chromothripsis	11
Figure 5. Mitotic errors and chromosome tethering in genomic instability	13
Figure 6. The distribution of ecDNAs in the cells	16
Figure 7. Schematic overview of the model system	21
Figure 8. Growth curves of fibroblasts from all four patients	22
Figure 9. Cellular morphology and DAPI staining of LFS primary fibroblasts across different passages.....	23
Figure 10. Phenotypic profiling identifies increased frequencies of chromatin bridges and fold-back inversions already at early and crisis passages.....	25
Figure 11. Dynamics of micronuclei formation across passages in LFS cells.....	26
Figure 12. Mitotic defects in LFS cells with increased multipolar spindle formation during crisis	27
Figure 13. SV profiles across passages in LFS041 using Strand-seq.....	29
Figure 14. Clustering on SVs and pairwise distance distributions in Strand-seq data	32
Figure 15. Single-cell Strand-seq and bulk WGS analysis of complex rearrangements in LFS041 cells	35
Figure 16. Ploidy assessment of metaphase spreads for patients LFS041 and LFS087 shows polyploidization in post-crisis and late passages as compared to early passage cells	36
Figure 17. UMAP visualization from HIPSD-seq (scDNA seq) showing clustering of cells across different passages of LFS041	38
Figure 18. Heatmap representation of copy number profiles in LFS041 samples analysed using HIPSD-seq	39
Figure 19. ReConPlots of SVs and somatic CNAs identified from WGS data for patient LFS041 p.63 (post-crisis).....	41
Figure 20. ReConPlots of SVs and somatic CNAs identified from WGS data for patient LFS087 p.195 (late)	42
Figure 21. FISH-based analysis of TERT and ALT telomere maintenance mechanisms in LFS	44
Figure 22. TERT-FISH analysis shows gains and losses in TERT locus in LFS172 post-crisis passage (p.72)	45
Figure 23. Morphological and growth characteristics of LFS <i>hTERT</i> ⁺ and negative control cells	46
Figure 24. Quantifications of Chromatin bridges and micronuclei in LFS174	47
Figure 25. Fraction of the genome altered per cell in LFS174 transduced cells using PTA ...	48
Figure 26. Copy number analysis for LFS174 hTERT and pMX-GFP control cells at passage 31 from PTA sequencing	49
Figure 27. Quantification and analysis of ecDNA in LFS041 and LFS087 Cells.....	50
Figure 28. High-confidence gene fusions identified from bulk RNA-seq.....	51
Figure 29. Representative example of the RFX8-TRIM40 gene fusion in LFS087 p.195 (late passage), identified with bulk RNA-seq	52
Figure 30. Differentially regulated proteins across cellular passages revealed by mass spectrometry-based proteomics	54

Figure 31. GSEA enrichment analysis of gene ontology (GO) in LFS041 fibroblasts using mass spectrometry-based proteomics	56
Figure 32. GSEA Reactome and KEGG pathway analysis in LFS041 proteomics data	58
Figure 33. Differential protein expression of selected pathways and biological processes based on GSEA GO proteomics data analysis in LFS041	60
Figure 34. Transcription factor enrichment co-expression network	61
Figure 35. Volcano plots of differentially expressed genes (DEGs) across cellular passages from bulk RNA-seq of LFS041	62
Figure 36. GSEA GO analysis in LFS041 bulk RNA-seq data.....	64
Figure 37. GSEA Reactome and KEGG pathway analysis in LFS041 bulk RNA-seq data....	66
Figure 38. Differential gene expression of selected pathways and biological processes based on bulk RNAseq analysis in LFS041.....	68
Figure 39. Impact of <i>TP53</i> / <i>MDM2</i> ⁺ on hypertranscription (reanalysis of TCGA data)	71
Figure 40. Graphical abstract summarising the main findings.....	87
Supplementary Figure 1. Phenotypic characteristics of LFS skin-derived fibroblasts.....	116
Supplementary Figure 2. SV profiles across passages in LFS041 using Strand-seq	117
Supplementary Figure 3. Representative Strand-seq examples of chromosomes 5 and 8 in early passage cells (LFS041 p.22)	118
Supplementary Figure 4. Strand-seq pseudobulk analysis of LFS041	118
Supplementary Figure 5. Heatmap representation of copy number profiles in LFS087 analysed using HIPSD-seq.....	119
Supplementary Figure 6. Circos plots depicting somatic alterations identified from WGS data of LFS041 p.63 (post-crisis) using OncoAnalyser	Error! Bookmark not defined.
Supplementary Figure 7. Circos plots depicting somatic alterations identified from WGS data of LFS087 p.195 (late) using OncoAnalyser	121
Supplementary Figure 8. Log10 of cumulative population doubling levels (PDLs) in LFS172	122
Supplementary Figure 9. ecDNA structures identified in LFS041 (p.63) and LFS087 (p.195) using AmpliconArchitect analysis of bulk WGS data	122
Supplementary Figure 10. TRIM40 fusions and expression in LFS primary fibroblasts.....	123
Supplementary Figure 11. ERC1 fusions and expression in LFS primary fibroblasts	124
Supplementary Figure 12. COL5A1 fusions and expression in LFS primary fibroblasts.....	125
Supplementary Figure 13. GSEA GO enrichment network visualizing interconnected pathways in of LFS041 proteomics data.....	127
Supplementary Figure 14. GSEA Reactome enrichment network visualizing interconnected pathways in of LFS041 proteomics data.....	129
Supplementary Figure 15. Volcano plots of DEGs across cellular passages from bulk RNA-seq of LFS087.....	130
Supplementary Figure 16. Complementary GSEA GO analysis in LFS041 bulk RNA-seq data	130
Supplementary Figure 17. Complementary GSEA Reactome and KEGG pathway analysis in LFS041 bulk RNA-seq data	131

List of tables

Table 1. LFS skin-derived fibroblasts used in the project.....	90
Table 2. Oligonucleotide primers and adaptors.....	90
Table 3. List of R packages used for data analysis.	91
Table 4. List of Python packages used for data analysis.....	91
Table 5. List of software tools.....	92
Table 6. List of databases and online tools.	92
Table 7. Composition of MS lysis mix.....	112
Table 8. Composition of the amplification reaction mix.....	112

Abbreviations

ALT	alternative lengthening of telomeres
AML	acute myeloid leukaemia
BER	base excision repair
BFB	breakage-fusion-bridge
BNHL	B cell non-Hodgkin lymphoma
BRCA	breast invasive carcinoma
BrdU	bromodeoxyuridine
C	Crick
CDKN2A	cyclin dependent kinase inhibitor 2A
CESC	cervical squamous cell carcinoma and endocervical adenocarcinoma
ChEA3	ChIP-X Enrichment Analysis Version 3
CGH	comparative genomic hybridization
chr. bridges	chromatin bridges
ChRCC	chromophobe renal cell carcinoma
CNA	copy number alteration
CIN	chromosomal instability
CLL	chronic lymphocytic leukaemia
CNS	central nervous system
DCIS	ductal carcinoma <i>in situ</i>
DEG	differentially expressed genes
Del	deletion
DMF	dimethylformamide
DSB	double-strand break
dup	duplication
EBP1	ErbB3-binding protein 1
ecDNA	extrachromosomal circular DNA
EMBL	European Molecular Biology Laboratory
FACS	fluorescence-activated cell sorting
FCS	foetal calf serum
FDR	false discovery rate
FISH	fluorescence <i>in situ</i> hybridization
FSC-A	forward scatter – area
GDC	genomic data commons
gDNA	genomic DNA
G6PD	glucose-6-phosphate dehydrogenase
GIST	gastrointestinal stromal tumour
GO	gene ontology
GSEA	gene set enrichment analysis
GWAS	genome wide association studies
h1	homolog 1
H2	homolog 2
h2hINV	head-to-head inversion
HCC	hepatocellular carcinoma
HDAC	histone deacetylase
HDR	homology directed repair
HIPSD-seq	High-throughPut Single-cell Dna sequencing
HIPSD&R-seq	High-throughPut Single-cell Dna and Rna-sequencing
hom	homologous
HR	homologous recombination
hytx_fch	hypertranscription fold-change

idup	inverted duplication
inv	inversion
IQR	interquartile range
KEGG	Kyoto Encyclopedia of Genes and Genomes
KIRC	kidney renal clear cell carcinoma
KNN	K-nearest neighbours
Lagging chr.	lagging chromosomes
LFS	Li-Fraumeni Syndrome
LOH	loss of heterozygosity
Mb	megabase
mBAND	multiple banding
Mbp	megabase pairs
MDS	myelodysplastic syndrome
mFISH	Multicolour FISH
min	minute
MMR	mismatch repair
MPN	myeloproliferative neoplasm
MPNST	malignant peripheral nerve sheath tumour
MSI	microsatellite instability
NER	nucleotide excision repair
NIH	National Institutes of Health
NHEJ	non-homologous end joining
NOS	not otherwise specified
ORC	origin recognition complex
p.	passage
PARP	poly-ADP ribose polymerase
PCPG	pheochromocytoma and paraganglioma
PDL	population doubling level
PDX	patient-derived xenograft
PEI	polyethylenimine
PGLS	6-phosphogluconolactonase
pH3	phospho-histone H3
PPAR	peroxisome proliferator-activated receptor
PRAD	prostate adenocarcinoma
PTA	primary template-directed amplification
RCC	renal cell carcinoma
RNA-seq	RNA sequencing
ROS	reactive oxygen species
SCC	squamous cell carcinoma
scDNA	single-cell DNA
SDS	Sodium dodecyl sulfate
sec	second
SEM	standard error of the mean
SHH	sonic hedgehog
SKCM	skin cutaneous melanoma
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SSC-A	side scatter – area
SSC-W	side scatter – width
STES	stomach and oesophageal carcinoma
SV	structural variant
t2tINV	tail-to-tail inversion
TCA	tricarboxylic acid
TCC	transitional cell carcinoma

TCEP	tris(2-carboxyethyl)phosphine
TERT	telomerase reverse transcriptase
TCGA	The Cancer Genome Atlas
TFEA	transcription factor enrichment analysis
UCEC	uterine corpus endometrial carcinoma
UMAP	uniform manifold approximation and projection
UV	ultraviolet
vs.	Versus
W	Watson
WGS	whole genome sequencing
WT	wildtype

1. Introduction

1.1. Genomic instability and cancer development

Genomic instability refers to the increased tendency of a cell to accumulate DNA mutations and chromosome rearrangements¹⁻³. This process drives the appearance of malignant cells and promotes cancer progression¹⁻³. Genomic instability is a major hallmark of cancer, detected in the vast majority of cancers¹⁻³.

Ranging from small-scale mutations to large-scale chromosomal rearrangements, genomic instability encompasses three main types²⁻⁴. The first type involves an increased number of nucleotide and base pair mutations, resulting in nucleotide-scale alterations^{3,4}. These mutations can arise due to defects in genes involved in two DNA repair machineries, namely base excision repair (BER) and nucleotide excision repair (NER), as well as from errors occurring during replication^{3,4}. The second type of genomic instability is linked to mutations in mismatch repair (MMR) genes, leading to microsatellite instability (MSI)³⁻⁵. Finally, chromosomal instability (CIN), the third and most common type, results in structural and numerical chromosomal abnormalities, including aneuploidy³⁻⁵.

Genomic instability arises from a complex interplay of various cellular processes. A key contributor includes defects in DNA damage checkpoints, involving genes such as *TP53* and *ATM*, which play crucial roles in halting the cell cycle to allow for DNA repair³. Mutations in caretaker genes further undermine the integrity of DNA repair pathways^{1,3,6}. These repair pathways include BER, NER, MMR, and double-strand break (DSB) repair mechanisms such as homologous recombination (HR) and non-homologous end joining (NHEJ); inactivation of these pathways significantly contributes to genomic instability³. Moreover, mitotic errors, such as chromosomal missegregation and defects in spindle checkpoint genes, as well as dysfunction in telomere maintenance, further exacerbate genomic instability^{1,3,4,7-10}. Additionally, replication stress, particularly at common fragile sites¹⁰⁻¹², transcription-replication conflicts¹³, epigenetic alterations^{3,14} and environmental or exogenous factors like ultraviolet radiation (UV)^{5,14} are important contributors to genomic instability.

Genomic instability has been shown to suppress immunity by enabling immune evasion¹⁵. However, without genomic instability, the low mutation frequency is insufficient to provide a selective advantage for cells, making tumour initiation unlikely¹⁵⁻¹⁷. Conversely, excessive genomic instability can lead to apoptosis or to pro-inflammatory signalling that enhances tumour cell immunogenicity by generating neoantigens, making cancer cells more susceptible to immune recognition¹⁵⁻¹⁷.

Introduction

Causes of genomic instability can stem from inherited mutator phenotypes or from sporadic, oncogene-induced replication stress events^{2,3,5}. These models underscore the dual role of genomic instability in cancer progression, acting as both a driver preceding other cancer hallmarks and facilitating their development, as well as a consequence of tumourigenesis, emerging as a downstream effect of oncogenic transformation^{1–3,5}.

1.2. The definition of chromothripsis

Cancer development is typically described as a stepwise process, beginning with a single mutation in a cell, which gradually accumulates additional mutations over time¹⁸. These mutations, along with epigenetic modifications, give the cells the ability to become malignant¹⁸. However, with recent advancements in sequencing technologies, a new form of genome instability has been discovered, termed chromothripsis, which refers to chromosome shattering¹⁹. This phenomenon has been characterized as a single catastrophic event¹⁹. During this process, up to hundreds of DNA DSBs occur in one or few chromosomes, which are shattered and subsequently repaired by error-prone DNA repair mechanisms, resulting in extensive chromosome rearrangements that can promote cancer^{19–21}.

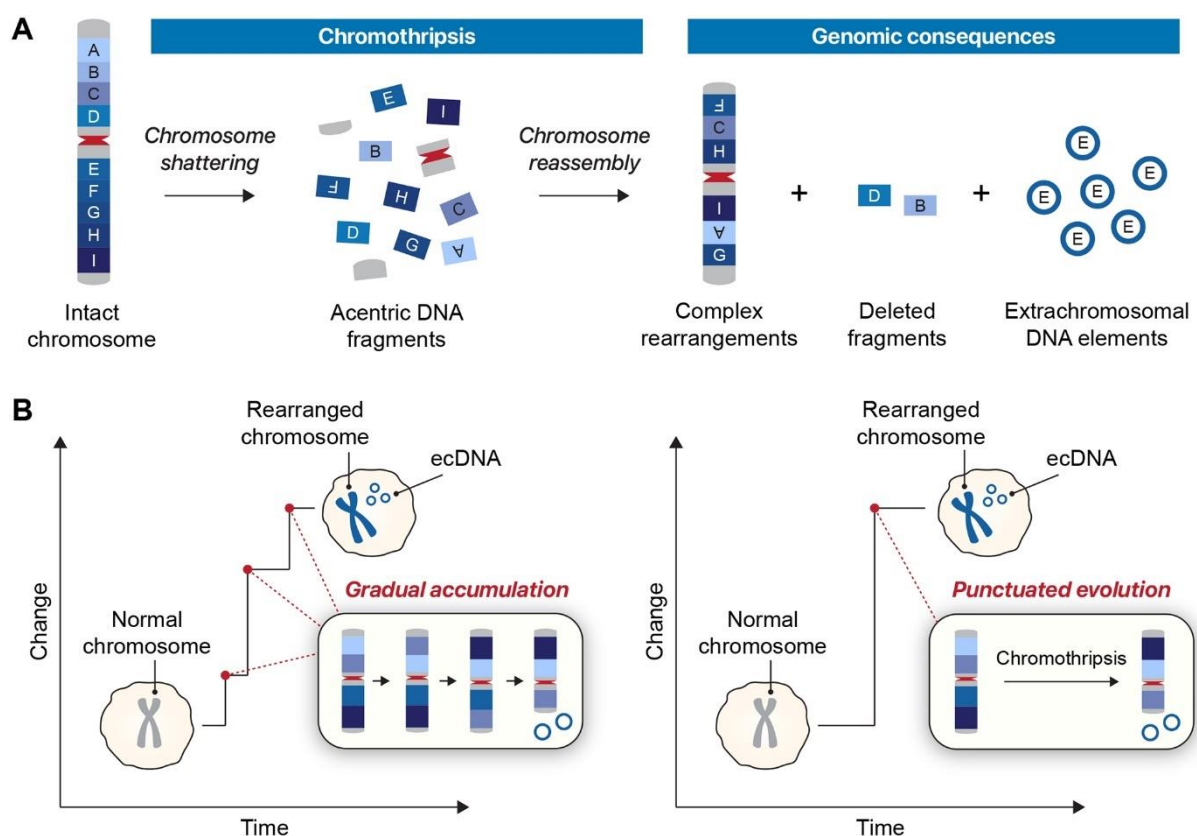


Figure 1. Chromothripsis and the evolutionary consequences of genomic chaos. A. Chromosome shattering and reassembly, resulting in a derivative chromosome harbouring massive chromosomal rearrangements, fragments loss, and the formation of extrachromosomal circular DNA (ecDNAs). **B.** The stepwise process of cancer development through gradual accumulation of mutations compared to the punctuated model, where sudden large-scale chromosomal rearrangements rapidly reshape the genome. Figure taken under permission from Mazzagatti et al. (2024)²¹.

One of the key outcomes of chromothripsis includes the formation of derivative chromosomes, which result from the rearranged segments being stitched together in a seemingly random order and orientation (Figure 1A)^{19–21}. These derivative chromosomes harbour numerous complex chromosomal rearrangements, including copy number alterations (CNAs), inversions and fusions^{19–21}. Furthermore, chromothripsis causes loss of chromosomal fragments, which can lead to the deletion of key genes, including tumour suppressor genes^{19–21}. The process can also give rise to extrachromosomal circular DNA (ecDNA), which may harbour amplified oncogenes^{19–21}. Notably, the rapid and massive changes in the genome resulting from chromothripsis align with the punctuated model of evolution, where extensive genomic rearrangements occur in a rapid manner rather than through continuous progressive changes (Figure 1B)^{18,20–22}.

1.3. Chromothripsis prevalence across human cancers

Chromothripsis was first described by Stephens et al. in 2011¹⁹. An unusual pattern of clustered genomic rearrangements was identified in the genome of a patient with chronic lymphocytic leukaemia (CLL)¹⁹. Although initially considered a rare event, affecting only 2–3% of cancers¹⁹, subsequent research has revealed a significantly higher prevalence, with chromothripsis detected in 30–50% of all cancer cases^{23–25}. However, reported frequencies vary widely across studies due to differences in detection methods, criteria for defining chromothripsis, as well as variations in sample representation, including cancer type, disease stage, and cohort size^{20,24}.

The prevalence of chromothripsis is highly cancer-type dependent^{20,24}. Some malignancies, such as liposarcomas, exhibit chromothripsis in nearly all cases²³. Other cancers, including osteosarcoma, melanoma, breast adenocarcinoma, and liver adenocarcinoma, show chromothripsis rates exceeding 60%^{23–25}. In contrast, certain cancer types, such as ovarian adenocarcinoma, demonstrate relatively low frequencies^{23–25}. Notably, no cases of chromothripsis have been identified in specific malignancies, including cervical adenocarcinoma, myelodysplastic syndrome (MDS), soft-tissue Ewing sarcoma and neuroblastoma^{23–25}. Furthermore, chromothripsis prevalence can vary drastically even within a single organ or tissue^{20,23–25}. For instance, in the nervous system, malignant peripheral nerve sheath tumours (MPNSTs) exhibit chromothripsis in nearly 100% of the reported cases, whereas central nervous system pilocytic astrocytomas show no prevalence, highlighting the heterogeneity and complexity of this phenomenon^{20,23,24}.

Despite the complex and stochastic nature of chromothripsis, emerging evidence suggests that this phenomenon is detected more frequently on certain chromosomes, with the specific pattern of involvement varying across cancer types^{23,26–28}. For instance, chromothripsis events are frequently observed on chromosomes 3 and 5 in kidney renal cell carcinoma, whereas alterations on chromosome 12 are more commonly associated with AML and liposarcomas^{23,26}.

Introduction

Similarly, breast cancer has shown recurrent chromothripsis on chromosomes 11 and 17²⁷, while chromosomes 5, 12, and 17 have been found to be enriched in osteosarcoma²⁸. Notably, chromosome 17 has been found to be affected in multiple cancers, likely because it harbours the tumour suppressor gene *TP53*^{23,29–32}. A strong association exists between chromothripsis and *TP53* loss of function, with tumours harbouring *TP53* mutations showing over 50% higher odds of undergoing chromothripsis compared to *TP53* wildtype (WT) cells^{23,29–33}.

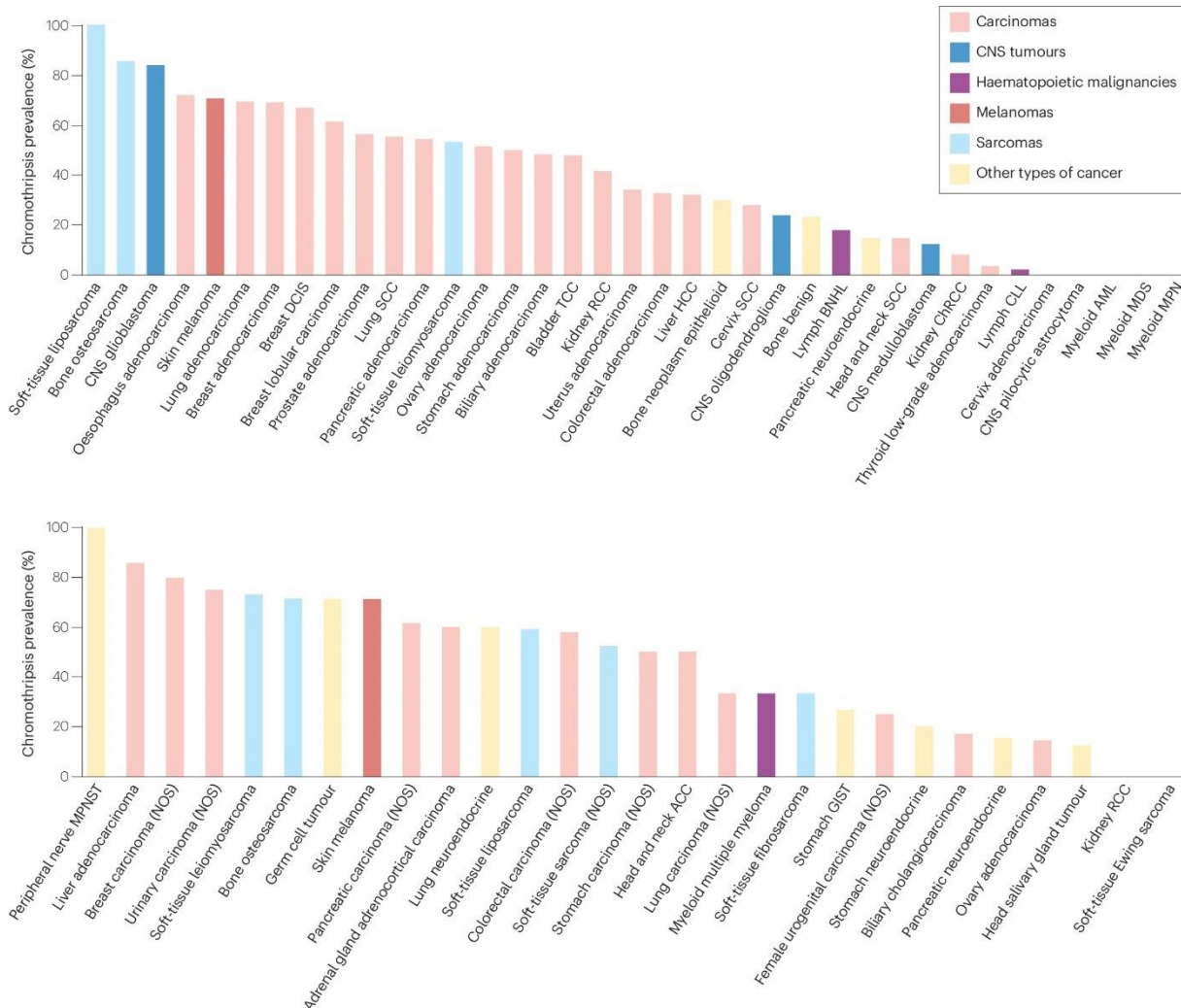


Figure 2. Chromothripsis frequency across human cancers. The upper panel presents data from Cortés-Ciriano et al. (2020), showing chromothripsis prevalence in 37 cancer types based on whole-genome sequencing analysis of 2,658 tumours. The lower panel displays results from Voronina et al. (2020), depicting chromothripsis prevalence in 28 tumour types encompassing common adult cancers, based on an analysis of 634 tumours. ACC: adenoid cystic carcinoma, AML: acute myeloid leukaemia, BNHL: B cell non-Hodgkin lymphoma, ChRCC: chromophobe renal cell carcinoma, CLL: chronic lymphocytic leukaemia, CNS: central nervous system, DCIS: ductal carcinoma in situ, GIST: gastrointestinal stromal tumour, HCC: hepatocellular carcinoma, MDS: myelodysplastic syndrome, MPN: myeloproliferative neoplasm, MPNST: malignant peripheral nerve sheath tumour, NOS: not otherwise specified, RCC: renal cell carcinoma, SCC: squamous cell carcinoma, TCC: transitional cell carcinoma. Figure adapted from Simovic-Lorenz and Ernst (2024), under the terms of the Creative Commons CC BY license²¹.

However, the exact mechanisms underlying this relationship remain incompletely understood²³. Nearly 60% of cancers exhibiting chromothripsis do not harbour *TP53* inactivating mutations or *MDM2* amplifications, as *MDM2* is a negative regulator of *TP53* and acts as an antagonist in the *TP53* pathway^{23,34,35}. Furthermore, mutations in other genes, such as *ATM* and those involved in DSB repair machinery like *BRCA1* and *BRCA2*, have also been linked to chromothripsis^{24,36}.

1.4. Characteristics of chromothripsis

Chromothripsis is part of a broader category of genomic catastrophes collectively known as chromoanagenesis, which includes other forms of large-scale genomic rearrangements, namely chromoanasythesis and chromoplexy^{30,37,38}. While these events all lead to massive genomic restructuring, they differ in their underlying mechanisms and genomic signatures^{30,37,38}.

In chromoanasythesis, the complex chromosomal rearrangements arise due replication-based errors, leading mainly to copy number gains, including segmental duplications and triplications, as well as templated insertions^{37,38}. In addition to its role in certain cancers, chromoanasythesis has also been implicated in congenital disorders³⁹. Chromoplexy, another form of chromoanagenesis, is characterized by complex chained rearrangements involving up to eight chromosomes^{37,38}. These events typically generate multiple fusions and translocations while maintaining a relatively balanced genome with minimal CNAs^{37,38}.

In contrast, chromothripsis involves a markedly higher number of DSBs and a greater degree of deletions and fragment losses as compared to the other two types of chromoanagenesis^{30,37,38}. A defining feature of chromothripsis is the clustered copy number oscillations, usually alternating between two or three copy number states (canonical chromothripsis)^{24,37,38}. However, a balanced form of chromothripsis, which lacks significant copy number changes, has also been reported⁴⁰. The resulting derivative chromosome often harbours complex structural rearrangements, including inversions, translocations, and gene fusions, which contribute to oncogene activation and tumour suppressor loss^{20,21,37,41,42}. A hallmark of chromothripsis is the formation of ecDNA, which can amplify oncogenes and promote tumour progression^{20,38}.

Beyond cancer, chromothripsis has been observed in congenital disorders^{37,38}. In a notable case, McDermott et al. reported an instance in which chromothripsis led to the spontaneous cure of WHIM syndrome, an immunodeficiency disorder⁴³. Importantly, chromothripsis has not only been identified in humans but across various species, including plants and rodents³⁸.

1.5. Key criteria for identifying chromothripsis

The criteria used to identify chromothripsis have evolved since its initial discovery^{19,23,44}. Initially, Stephens et al. revealed six key features for calling chromothripsis, which is mainly based on whole genome sequencing (WGS) as a gold standard¹⁹. These criteria included localized and complex rearrangement patterns, copy number changes alternating between two states, clustered breakpoints, dispersed junctions and interchromosomal rearrangements¹⁹.

Later, Korbelt and Campbell refined these criteria by introducing additional computational and statistical guidelines for identifying chromothripsis⁴⁴. The revised criteria involved 1) breakpoints clustering within a confined genomic region, 2) segmental copy number oscillations, 3) retention of heterozygosity in copy number-neutral regions, 4) haplotype-specific rearrangements, 5) randomized orientation and assembly of DNA segments and 6) the mapping of the derivative chromosome with alternating head-to-tail and tail-to-head connections between DNA segments⁴⁴.

As sequencing technologies advanced, additional methods for detecting chromothripsis were developed. Rausch et al. proposed a threshold-based approach for identifying chromothripsis through at least ten alterations across two or three copy number states on an individual chromosome³³. This method has been adopted in several studies^{30,45}. However, the minimum number of breakpoints per chromosome required to classify an event as chromothripsis varies between studies, with some requiring fewer than 10 and others more than 50 breakpoints^{30,45}.

More recently, computational tools have been developed to facilitate chromothripsis detection^{23,46,47}. ShatterSeek, for instance, builds upon the statistical criteria established by Korbelt and Campbell, incorporating additional features to identify complex oscillation patterns and interchromosomal structural variants (SVs)²³. In addition, ShatterSeek adapts the scoring criteria to account for aneuploidy and tumour heterogeneity by ignoring loss of heterozygosity (LOH) analysis in complex aneuploid samples, enhancing sensitivity in identifying chromothripsis events²³.

These computational tools and algorithmic approaches aim to standardize chromothripsis detection across studies and distinguish it from gradual accumulation of rearrangements^{23,44}. Despite these advancements, variations in detection methods and interpretation persist. Moreover, accurately detecting chromothripsis remains challenging due to the complexity and variability of genomic rearrangements, making it difficult to distinguish true chromothripsis events from other forms of complex genomic instability^{23,45}.

1.6. Methods to detect chromothripsis and CNAs

Various approaches have been developed to study chromothripsis and CIN in general. These methods often rely on a combination of experimental and computational approaches to capture the complex genomic alterations underlying chromothripsis. WGS has become the gold standard for identifying chromothripsis and CNAs^{19,30,31}. Through pair-end or mate-pair sequencing^{19,30}, together with computational tools, such as CTLP-Scanner⁴⁶, ShatterProof⁴⁷ or ShatterSeek²³, breakpoint locations and the orientation of genomic segments, including head-to-head and tail-to-tail inversions, can be accurately identified³⁰.

Array-based methods, such as comparative genomic hybridization (CGH), have also been employed for detecting CNAs, as well as single nucleotide polymorphism (SNP) arrays^{30,31,48}. However, these methods are limited in detecting balanced chromothripsis, as inversions remain undetectable³⁰.

Cytogenetic and molecular techniques like fluorescence *in situ* hybridization (FISH) have been also used in chromothriptic studies to detect or validate specific rearrangements^{30,31,48}. Multicolour FISH (mFISH) and multiple banding (mBAND) techniques are effective for detecting translocations and inversions at the chromosome or chromosome arm level^{30,31,48}. Additionally, FISH is valuable for identifying ecDNAs^{30,49}.

However, the field is increasingly moving towards genomic techniques, with both low and high-throughput sequencing technologies at the single-cell and bulk level offering enhanced precision and sensitivity of detection⁵⁰. Single-cell DNA (scDNA) sequencing has emerged as a powerful tool for studying chromothripsis, particularly in cancer, where chromothripsis may occur in a subpopulation of cells, contributing to intratumoural heterogeneity^{50–52}. These approaches can reveal clonal evolution of chromothripsis events and provide insights into their functional consequences^{51,52}.

One interesting scDNA sequencing technique is strand-specific sequencing (Strand-seq), which is based on labelling newly synthesized DNA strands with bromodeoxyuridine (BrdU) incorporation during DNA replication (Figure 3A)^{53,54}. Following replication, the BrdU-labeled strands are selectively removed, allowing only the parental template strands to be sequenced^{53,54}. This approach preserves strand-specific information, enabling the detection of SVs, inversions, sister chromatid exchanges, translocations, and complex chromosomal rearrangements, including chromothripsis, with high resolution^{53–55}. Additionally, Strand-seq allows for the precise mapping of genomic rearrangements to individual haplotypes, facilitating the identification of fold-back inversions and breakage-fusion-bridge (BFB) cycles^{53–55}.

Introduction

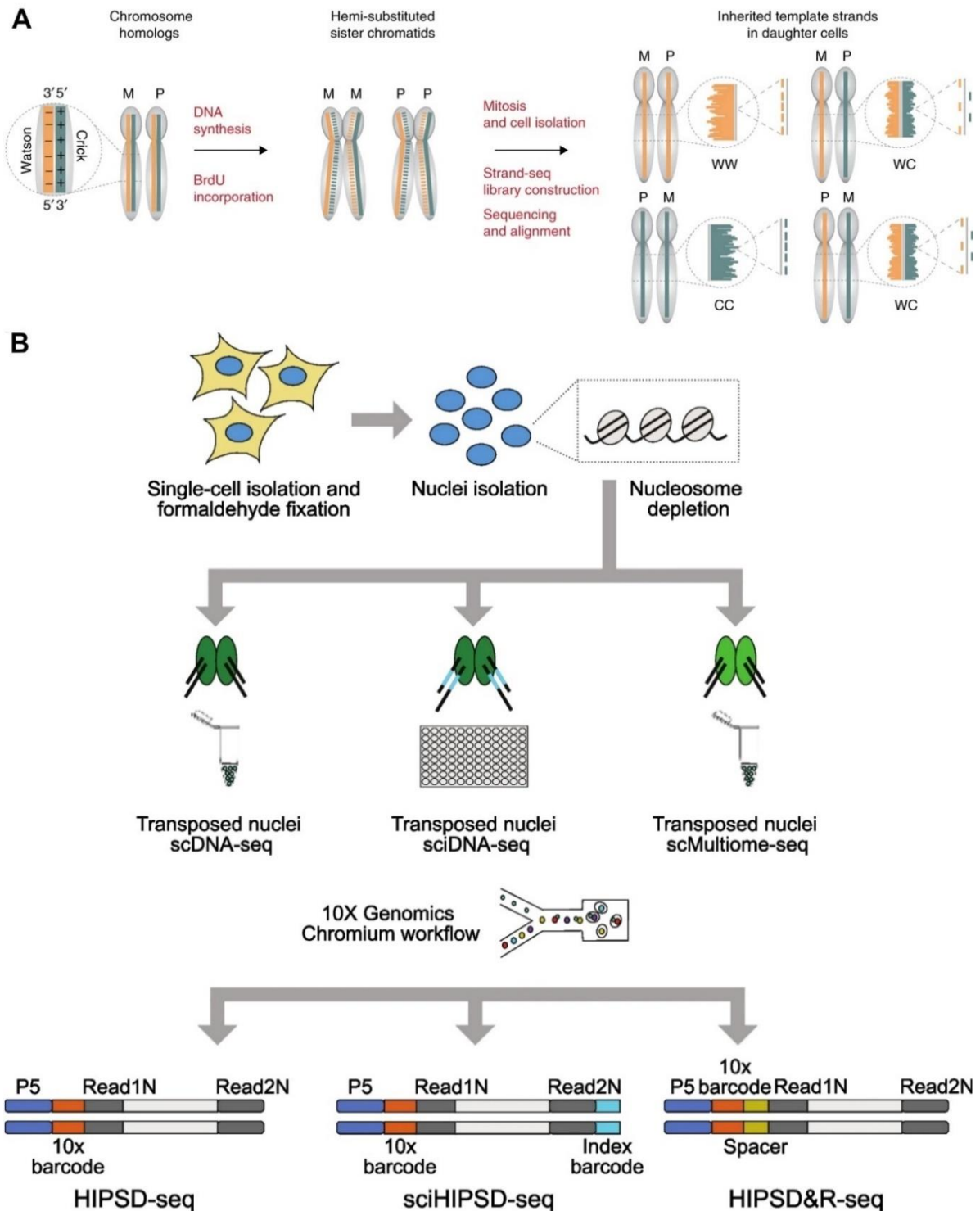


Figure 3. scDNA sequencing by Strand-seq and HIPSD&R-seq to detect and study genomic instability and chromothripsis. **A.** BrdU is incorporated into newly synthesized maternal (M) and paternal (P) chromosomes (dashed lines) during DNA replication. After one round of cell division and BrdU removal, each daughter cell retains only the original template strands, resulting in WW, CC, or WC configurations. Figure taken under permission from Sanders et al. (2017)⁵⁴. **B.** Schematic overview of the HIPSD&R-seq workflow. After cell fixation, nuclei isolation, and nucleosome depletion using sodium dodecyl sulfate (SDS), one of three protocols is applied for Tn5 tagmentation, library preparation, and sequencing, depending on the desired sequencing approach. HIPSD-seq (High-throughPut Single-cell Dna sequencing) is used for scDNA sequencing, HIPSD&R-seq for scDNA and scRNA multiome sequencing, and sciHIPSD-seq for ultra-high-throughput scDNA sequencing. Figure taken from Otoničar et al. (2024), under the terms of the Creative Commons CC BY license⁵².

High-throughput scDNA sequencing has revolutionized our ability to study genomic heterogeneity at the single-cell level, enabling detailed analyses of CNAs, clonal evolution, and cellular heterogeneity⁵⁰. These techniques overcome the limitations of bulk sequencing, which often masks heterogeneity within a population of cells^{50,52}. Many of these methods now offer multi-omics approaches, combining scRNA and scDNA sequencing⁵⁰. However, these methods often do not offer the capacity to sequence the entire genome, limiting their utility for detecting CNAs, chromosomal instability, and chromothripsis⁵². Several challenges are inherent to such methods, such as noise, amplification bias, low coverage, difficult bioinformatics analysis as well as high costs⁵⁰.

HIPSD&R-seq (High-throughPut Single-cell Dna and Rna-seq) is a recent multiomics method that allows sequencing of thousands of cells⁵². This method builds upon a modified version of the 10X Genomics platform, but includes a sodium dodecyl sulfate (SDS)-based nucleosome depletion and highly efficient Tn5 transposase for whole-genome sequencing (Figure 3B)⁵². Moreover, this method can be combined with combinatorial indexing, enabling ultra-high-throughput sequencing of tens of thousands of cells simultaneously (sciHIPSD-seq)⁵². Although its coverage may not match that of Strand-seq, HIPSD&R-seq provides significant benefits for both large-scale studies and in-depth analysis of complex genomic landscapes⁵².

In addition to these methods, long-read sequencing offers higher resolution for studying complex rearrangements and mapping breakpoints with greater accuracy compared to short-read sequencing^{56,57}. Moreover, spatial transcriptomics has emerged as a valuable tool in chromothripsis research⁵⁸. A recent study of Sonic hedgehog (SHH) medulloblastoma revealed greater intra-tumour heterogeneity in chromothriptic medulloblastomas, along with higher proliferation and the presence of cells with stem-like features compared to medulloblastomas without chromothripsis⁵⁸.

1.7. Mechanisms underlying chromothripsis

Several mechanisms have been proposed to explain the process of chromothripsis initiation, including telomere dysfunction, micronuclei formation, aberrant mitosis, replication stress, and external insults^{23,30,32,59}. These mechanisms are highly interconnected and each contributes to our understanding of this intricate process^{23,30,32,59}.

1.7.1. Telomere dysfunction and BFB cycles

Cells experience progressive telomere attrition after each cell division, and once the telomeres become critically short, they lose their ability to protect chromosomes from end-to-end fusions^{60–62}. This leads to the formation of dicentric chromosomes and genomic instability, a process known as telomere crisis^{59–62}. In response to telomere shortening, most cells enter replicative senescence or activate apoptosis to prevent cancer initiation^{37,62}. However, in

Introduction

certain cases, for instance when tumour suppressor genes such as *TP53* or *RB1* are defective, cells may bypass telomeric crisis and form telomeric fusions, resulting in chromosomes with two centromeres^{37,62}. During mitosis, each centromere attaches to opposite spindle poles, pulling the chromosomes in different directions^{37,62}. This process can lead to the formation of a chromatin bridge, which ultimately breaks, resulting in the re-fusion of the broken ends in daughter cells due to the absence of telomeres, thereby creating another dicentric chromosome⁶². This repeated cycle is termed BFB cycles (Figure 4A)^{37,62}.

The asymmetric and random nature of chromosome breakage during BFB cycles results in significant genomic alterations^{37,62}. For instance, one daughter cell may inherit a shorter chromosome that lacks crucial genetic material, leading to LOH, while the other cell retains a chromosome with a duplicated DNA fragment^{37,62}. Subsequent BFB cycles can result in inverted duplications and further genomic alterations^{37,62}.

Extensive research has aimed to elucidate the mechanistic basis of chromosome shattering associated with chromatin bridges^{63–65}. Maciejowski et al. suggested that these bridges often lead to nuclear envelope rupture, exposing DNA to TREX1-mediated fragmentation⁶³. However, this hypothesis has been challenged by studies indicating that inhibiting TREX1 activity or preventing nuclear membrane rupture does not eliminate or significantly delay BFB cycles, suggesting that additional mechanisms of breakage are at play⁶⁵. In 2020, Umbreit et al. found that mechanical forces generated by actomyosin contraction contribute significantly to the breakage of chromatin bridges⁶⁵. Thus, our current understanding posits that mechanical stretching primarily drives dicentric chromosome shattering during mitosis, while TREX1 may enhance this process (Figure 4B)^{59,62,63,65}. However, the complete molecular basis of this process is still to be explored^{20,59}.

Moreover, after the initial breakage of dicentric chromosomes, daughter cells often experience DNA replication stress on the fragmented chromosomes, leading to further genomic instability⁶⁵. As a result, by the second mitosis, these damaged chromosomes often missegregate and can become encapsulated into micronuclei⁶⁵.

1.7.2. Micronuclei formation

The formation of micronuclei has emerged as one of the most well-established and studied mechanisms contributing to chromothripsis and genomic instability^{30,37}. Micronuclei formation typically follows mitotic errors, such as defective kinetochore-spindle attachments, chromatin fragmentation, missegregation, and lagging chromosomes^{30,32,59,66}. These errors result in the encapsulation of whole chromosomes, acentric chromosome fragments, or sometimes more than one chromosome, into small, round, extra-nuclear structures known as micronuclei^{31,32,66}.

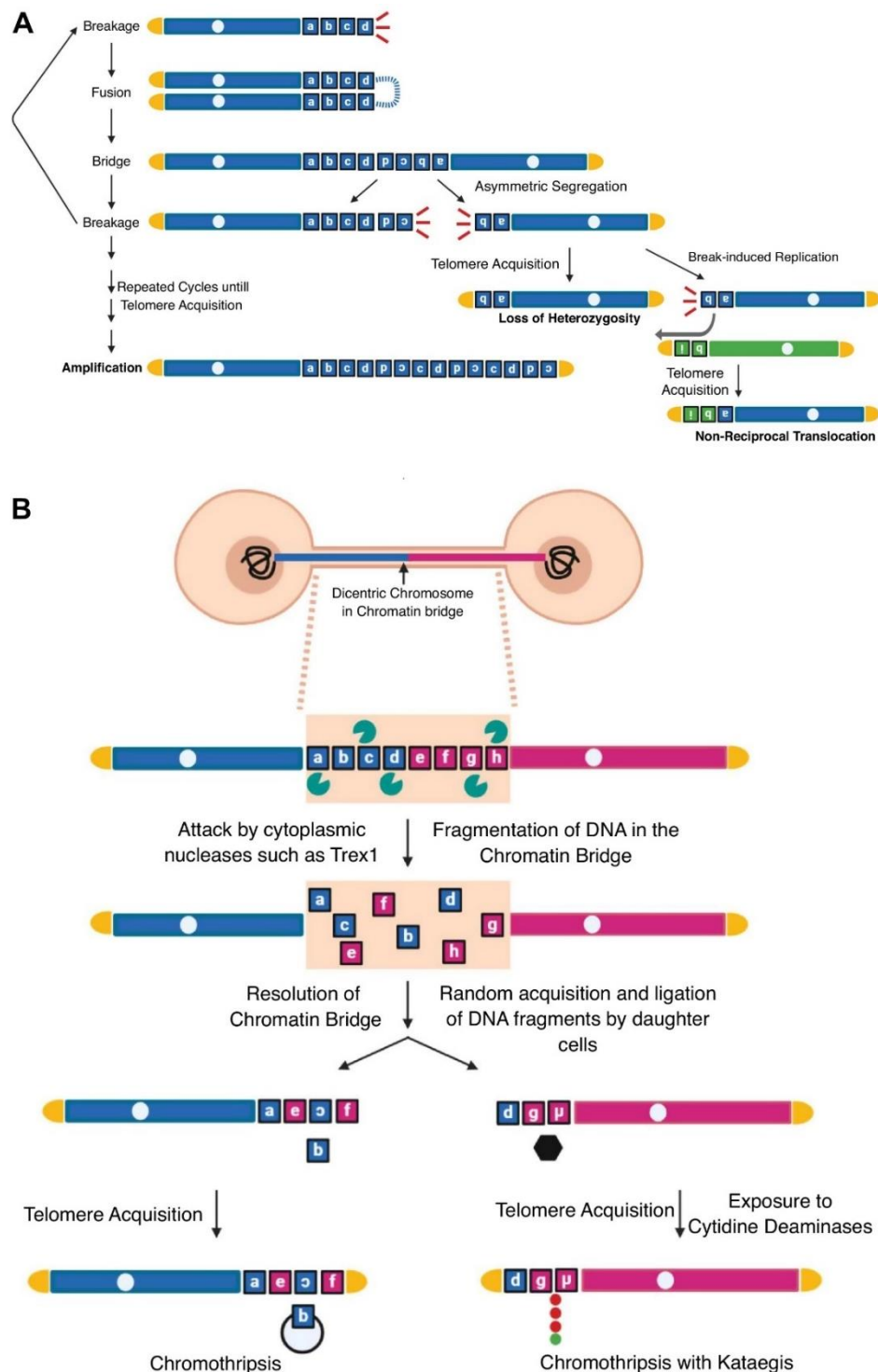


Figure 4. Mechanisms underlying BFB cycles and their role in chromothripsis. A. BFB cycle initiation and progression following telomere dysfunction. Telomere shortening leads to uncapped chromosomes that fuse, forming dicentric chromosomes. Asymmetric breakage during cell division results in various chromosomal rearrangements, including LOH, amplifications, and inversions. Broken ends may acquire new telomeres through telomerase activation or break-induced replication, with the latter potentially causing non-reciprocal translocations. Persistent telomere uncapping in daughter cells drives repeated BFB cycles in subsequent mitoses. Figure taken under permission from Bhargava et al. (2020)⁶². **B.** During mitosis, chromatin bridges resulting from dicentric chromosomes can be fragmented through two proposed mechanisms: mechanical breakage via actomyosin contraction and cleavage by the cytoplasmic exonuclease TREX1. These processes can lead to chromothripsis, which is occasionally accompanied by kataegis – a hypermutation pattern defined by clusters of closely spaced single-base pair substitutions^{62,67}. Figure adapted under permission from Bhargava et al. (2020)⁶².

Introduction

Chromosomes or chromosome fragments within micronuclei are highly prone to DNA damage, which is evidenced by the accumulation of γ H2AX, a well-established DNA damage marker^{32,68}. Micronuclear envelopes frequently rupture, often due to defective assembly of nuclear components, such as the nuclear lamina and nuclear pore complexes, in addition to mechanical forces from actin cytoskeletal contractions^{69–71}. This rupture leads to further DNA damage and allows endoplasmic reticulum invasion, potentially exacerbating the damage^{71,72}. Exposed micronuclear DNA activates the cytosolic DNA sensor cGAS, triggering innate immune responses via the cGAS-STING pathway^{72,73}. However, the endoplasmic reticulum-associated nuclease TREX1 can mitigate this process by degrading cytosolic DNA in ruptured micronuclei, limiting cGAS activation and subsequent inflammatory signalling⁷².

Research from David Pellman's lab and others revealed that DNA replication in micronuclei is delayed and often incomplete, as ruptured micronuclei lack critical components necessary for DNA replication and repair, such as replication licensing factors and DNA repair machineries^{68,71,74}. This results in slowed replication and stalled replication forks, leading to asynchronous replication with the main nucleus^{68,74}. Using Look-seq, an innovative technique combining live-cell imaging with scDNA sequencing, Zhang et al. observed that DNA within micronuclei undergoes extensive damage and fragmentation⁷⁴. Their scDNA sequencing analysis revealed that complex rearrangements and chromothripsis can occur within a single cell cycle, specifically within micronuclei⁷⁴.

Micronuclei can follow several paths during subsequent cell divisions^{32,74}. In some cases, micronuclei reintegrate into the main nucleus of a daughter cell, leading to the introduction of altered or fragmented DNA into the genome, which can drive genomic instability^{32,74}. In other instances, micronuclei remain separate from the main nuclei, being inherited by one daughter cell without merging with the primary nucleus^{32,74}. Alternatively, micronuclear DNA may degrade, leading to the loss of genetic material^{32,74}. Despite these significant advancements, many aspects of micronuclear dynamics remain poorly understood, and further research is required to fully elucidate the underlying mechanisms of chromothripsis and genomic instability associated with micronuclei³⁷.

1.7.3. Aberrant mitosis

Mitotic defects are a major source of both numerical and structural chromosomal abnormalities, often resulting in chromatin bridges and micronuclei formation^{21,29,65,74}. The upstream causes of these errors vary, ranging from centromere defects to spindle attachment issues, which can result in chromosome missegregation and lagging chromosomes^{21,29}. Cell cycle checkpoints act as crucial safeguards, monitoring the progression of cell division and halting the process when errors are detected^{21,29}. Each checkpoint plays a unique role in

maintaining genomic integrity, and failure at any of these checkpoints allows cells with chromosomal abnormalities to continue dividing, leading to genomic instability and large-scale chromosomal rearrangements, including chromothripsis^{21,29}. While these checkpoints are essential, it is important to note that DSB repair by NHEJ and HR is largely inactivated during mitosis²¹.

The spindle assembly checkpoint ensures proper kinetochore-microtubule attachments before allowing anaphase onset²¹. Defects in this checkpoint can lead to attachment errors between centromeres and microtubules, potentially resulting in chromosome missegregation²¹. The abscission checkpoint operates during the final stages of cytokinesis, delaying cell separation in the presence of chromatin bridges^{21,75}. This delay provides time for error correction and prevents the formation of aneuploid daughter cells^{21,75}. The DNA damage response checkpoint halts the cell cycle when DNA damage is detected, allowing time for repair mechanisms to address the issue²¹. In addition to its involvement in the process leading to chromosome breakage, mitosis is crucial in ensuring that fragmented chromosomes are properly segregated during cell division^{21,29}. To prevent the loss of the acentric genomic fragmented segments, the CIP2A-TOPBP1 complex plays a pivotal role in physically linking these fragments, ensuring that they are inherited by one of the daughter cells (Figure 5)^{21,29}. This tethering action relies on the recruitment of TOPBP1 to DNA damage sites^{21,29}.

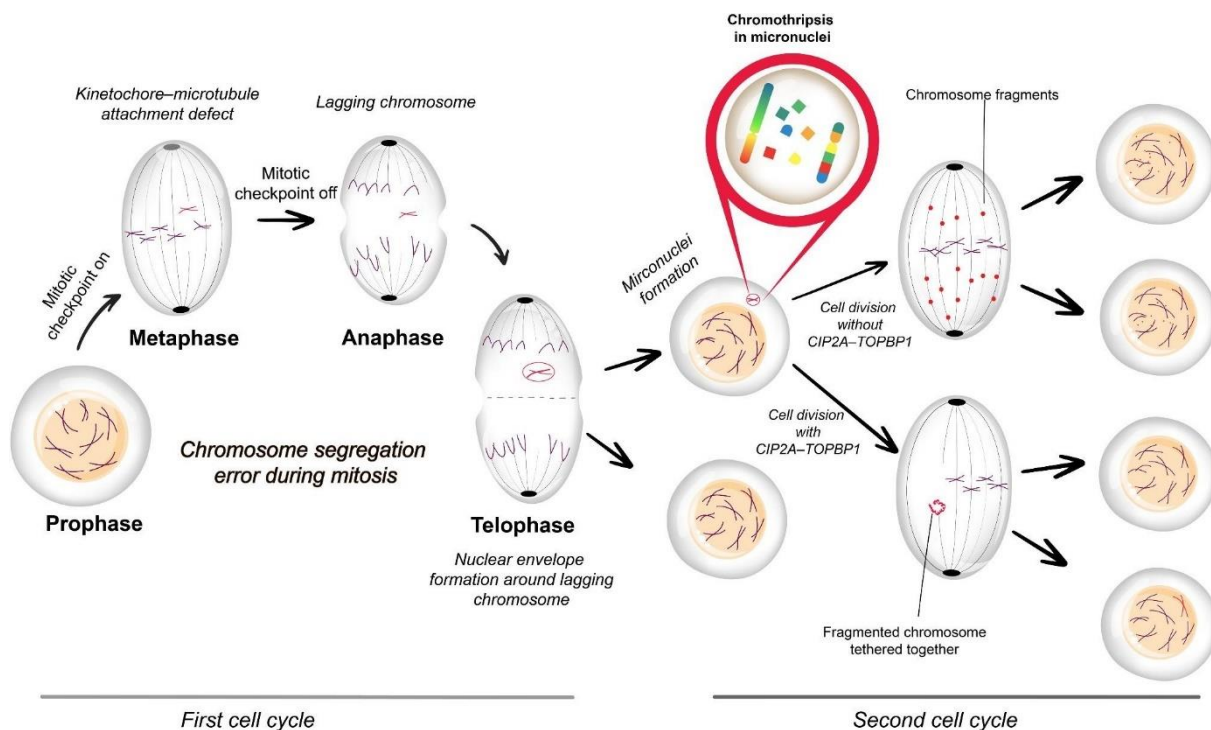


Figure 5. Mitotic errors and chromosome tethering in genomic instability. Errors at different stages of mitosis lead to chromosome missegregation, resulting in the formation of lagging chromosomes and micronuclei. In the second cell cycle, if chromosome binding mediated by the CIP2A-TOPBP1 complex occurs, fragmented chromosomes can remain connected, allowing for their co-segregation. However, without tethering, chromosome fragments exhibit a more dispersed distribution, which results in deleted fragments and elevated genomic instability. Image taken under permission from Ejaz et al. (2024)²⁹.

Introduction

By bridging broken chromosomes during cell division, CIP2A-TOPBP1 suppresses the formation of micronuclei and facilitates the segregation of chromosome fragments^{21,29}. Disruptions in this tethering process can lead to improper chromosome segregation, contributing to genomic instability and potentially promoting cancer progression^{21,29}.

1.7.4. Replication stress

DNA replication plays a critical role in driving chromothripsis, largely due to the extensive damage that it generates, which can lead to chromosome fragmentation^{76,77}. Replication stress, a condition where DNA replication forks stall or collapse, is triggered by factors such as the activation of oncogenes, depletion of nucleotides, conflicts between transcription and replication, or DNA damage^{12,77–79}. Ongoing replication stress can lead to the formation of DNA DSBs when the replication forks collapse, potentially leading to large-scale CIN^{77,78,80}.

In cells with defective repair pathways, DNA repair mechanisms may not be able to cope with the excessive DNA damage^{78,79}. This can force cells to enter mitosis with under-replicated DNA, potentially triggering premature chromosome condensation^{76,79,81}. These events may result in chromosome fragmentation and micronuclei formation^{12,76,79}. Severe replication stress exacerbates this process by leaving certain genomic regions under-replicated, creating fragile sites that are highly prone to breakage during mitosis, potentially leading to chromothripsis^{12,77,79}. Oncogene-induced replication stress can further amplify these effects, exacerbating the cycle of replication stress and chromosomal damage^{79,82}.

Additionally, Wilhelm et al. showed that even moderate levels of replication stress can cause premature disengagement of the centrioles and result in multipolar spindle formation⁸³. This ultimately leads to chromosomal missegregation and micronuclei formation, as well as numerical chromosomal abnormalities⁸³. In line with this, Shaikh et al. reported that distinctive patterns of CNAs are caused due to replication errors and suggested that this process drives cancer evolution⁹¹.

1.7.5. Additional proposed mechanisms of chromothripsis

Beyond the major mechanisms of chromothripsis initiation, other mechanisms have been suggested to contribute to chromothripsis. One such mechanism is aborted apoptosis, where exposure to high levels of DNA DSBs, such as those induced by chemotherapy, typically triggers apoptosis^{32,59}. In cells with CIN that evade apoptosis, particularly those with *TP53* mutations, the accumulation of fragmented chromosomes may drive chromothripsis^{32,59}.

Furthermore, cells with defective DNA repair pathways, particularly those with impaired HR and NHEJ, have been shown to lead to complex chromosomal rearrangements and chromothripsis⁸⁴. Premature chromosome condensation has also been implicated, as cells

entering mitosis with incompletely replicated DNA are prone to chromosome pulverization and subsequently to chromothripsis^{32,59,74}. Ionizing radiation is another factor known to induce extensive DSBs, which can lead to micronuclei formation^{32,59}. Additionally, viral infections have been proposed to contribute to chromothripsis, although their exact role remains unclear and may involve additional cellular mechanisms⁵⁹.

1.8. Consequences of chromothripsis and selective advantages

Cells that undergo chromothripsis experience widespread DNA damage and complex chromosomal rearrangements^{20,21,32}. In most cases, these cells either undergo apoptosis or are eliminated through negative selection due to the loss of essential genes required for cellular function³². However, cells that survive this catastrophic event, particularly those harbouring defects in key tumour suppressor pathways – such as *TP53* – may acquire features conferring survival and proliferation advantages^{20,32,33}.

A hallmark consequence of chromothripsis is the formation of derivative chromosomes carrying numerous structural and numerical aberrations^{20,21,32}. The random reassembly of chromosomal fragments can result in clinically relevant gene fusions^{20,21,32,85,86}. For example, chromothripsis-driven fusions involving *PVT1* have been reported in medulloblastoma, contributing to tumour progression⁸⁵. Chromothripsis can also result in the loss of tumour suppressor genes^{20,23,32,86,87}. Studies have reported that chromothripsis can lead to the deletion or inactivation of essential genes such as *CDKN2A*, *MLH1*, *PTEN*, *BRCA1*, *BRCA2*, and *TP53*, thereby promoting genomic instability and increasing cancer susceptibility^{20,23,32,86}. In addition, chromothripsis can also drive oncogenesis through the amplification of oncogenes³². *MYC* amplification has been frequently observed as a result of chromothripsis in various cancers, leading to uncontrolled cell growth and tumour progression³².

Furthermore, telomere stabilization has been shown to play a crucial role in chromothriptic cells^{24,49}. Previous studies have shown that tumours with chromothripsis often exhibit distinct telomere maintenance mechanisms^{24,49}. Telomerase activation, typically through increased *TERT* (telomerase reverse transcriptase) expression due to the amplification of *TERT* locus, has been observed in chromothripsis-positive cases across various cancer types^{24,49}.

The second mechanism is the alternative lengthening of telomeres (ALT), which is a recombination-dependent process to maintain telomere length, often associated with mutations in *ATRX* or *DAXX* genes⁶². ALT has been implicated in some chromothriptic tumours, such as SHH medulloblastomas with *TP53* mutations⁴⁹. The activation of these telomere maintenance mechanisms is thought to stabilize the derivative chromosomes resulting from chromothripsis, preventing further genomic instability and allowing continued cell proliferation^{24,49}.

Introduction

One of the most significant consequences of chromothripsis is the ecDNA formation, a topic of current research interest due to its role in conferring selective advantages to cancer cells^{20,21,88}. ecDNA molecules range from tens of kilobases to several megabases in size, lacking centromeres and telomeres^{88,89}. ecDNA contributes to oncogene activation through various mechanisms, primarily via amplification of key oncogenes⁸⁸. Notable examples include *MYC*, *EGFR*, *MDM2*, *TERT*, *ERBB2*, and *CCND1*⁸⁸. Additionally, ecDNAs can lead to oncogenic expression through enhancer hijacking, where distal regulatory elements are brought into proximity with oncogenes, further driving their expression^{88,90}. The presence of ecDNAs has been linked to increased tumour heterogeneity and clonal evolution, thus contributing to therapy resistance and poor clinical outcomes^{88,91}. A study from Luebeck et al. linked ecDNAs with worse prognosis in Barrett's oesophagus⁹². Recently, Tang et al. have proposed a promising novel approach to target ecDNA-containing tumour cells by exploiting transcription-replication conflicts, highlighting potential novel therapeutic interventions⁹¹.

The distribution of ecDNAs during cell division is non-Mendelian due to the absence of centromeres⁹³. As a result, ecDNA is often unevenly segregated between daughter cells, contributing to intratumoural heterogeneity⁸⁸. Following mitosis, ecDNAs can reintegrate into chromosomal DNA, forming homogeneous staining regions (HSRs) – amplified chromosomal loci (Figure 6)⁸⁸. Alternatively, ecDNAs may become engulfed in micronuclei and eventually removed from the cell or persist as extrachromosomal elements⁸⁸. However, ecDNA formation, distribution, and inheritance are complex processes that are still being elucidated⁸⁸.

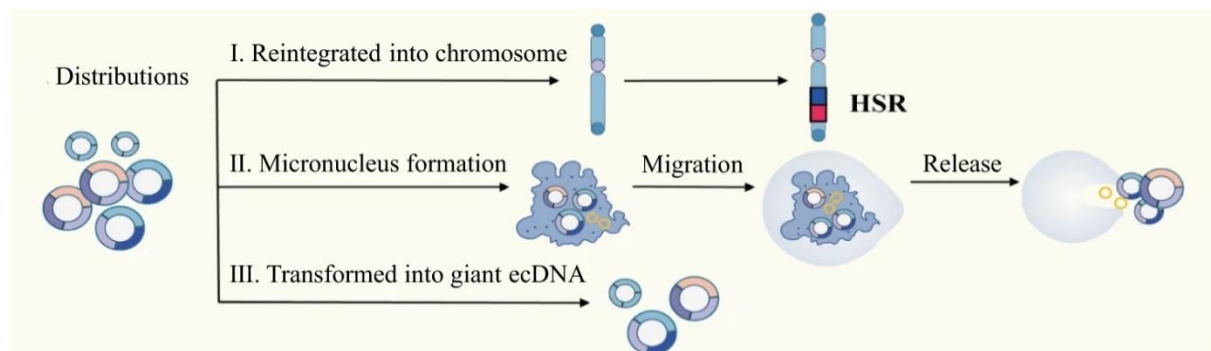


Figure 6. The distribution of ecDNAs in the cells. EcDNA can follow three main distribution paths during cell division: I) reintegrate into chromosomes, II) become encapsulated in micronuclei, potentially leading to their elimination or further amplification, or III) persist as extrachromosomal fragments. Figure adapted from Chen et al. (2023), under the terms of the Creative Commons CC BY license⁸⁸.

1.9. Clinical implications of chromothripsis

Chromothripsis has been associated with poor prognosis and shorter survival in multiple tumour types, such as neuroblastoma^{94,95}, medulloblastoma⁹⁶ and melanoma⁹⁷. Despite the aggressive nature of chromothripsis-driven tumours, the extensive DNA damage and complex genomic alterations that such tumours harbour present potential therapeutic vulnerabilities^{20,98}. One promising approach involves targeting DNA repair pathways, particularly in tumours with

HR deficiency, through the use of PARP (poly-ADP ribose polymerase) inhibitors^{58,96,99}. Previous studies from our lab demonstrated that combining PARP inhibitors with carbon ion therapy enhances tumour cell lethality in chromothriptic tumours, as HR-deficient cells exhibit increased sensitivity to radiation^{58,96,100}. Additionally, the synergistic use of PARP inhibitors with HDAC (histone deacetylase) inhibitors has been identified as a potential treatment strategy for chromothriptic tumours¹⁰¹. Other therapeutic strategies have also been explored, including targeting ecDNAs^{91,98}, disrupting the cGAS–STING signalling pathway¹⁰², and impairing microtubule–kinetochore attachment to promote chromosome missegregation and induce apoptosis through mitotic catastrophe¹⁰³.

1.10. Current models to study chromothripsis

To study chromothripsis, a variety of models to induce chromosomal rearrangements resembling this phenomenon have been developed. Current *in vitro* models primarily focus on inducing mitotic defects that lead to micronucleation and subsequent chromothripsis^{20,66,68,74,104,105}.

One of the pioneering models was developed in the lab of David Pellman, where nocodazole was used to disrupt microtubule polymerization, inducing mitotic arrest^{68,74}. This led to micronucleation and subsequently to chromothripsis^{68,74}. Building on this foundation, researchers have employed various antimitotic agents and genetic manipulations to induce chromothripsis-like events. These include inactivation of the Y chromosome centromere to cause centromere dysfunction^{66,104}, knockdown of *PICH* to create replication stress¹⁰⁵, and gene editing via CRISPR-Cas9 to generate targeted DNA breaks¹⁰⁶.

Other approaches have focused on inducing telomere crisis and BFB cycles, particularly through retroviral transduction of a dominant-negative *TRF2* allele^{63,64}. Additionally, researchers have employed chemical agents such as doxorubicin^{107,108}, paclitaxel¹⁰⁸, cisplatin¹⁰⁸, camptothecin¹⁰⁸, and methotrexate⁹⁸ to induce micronucleation through various mechanisms, including DNA damage and mitotic spindle defects. Viral infections, including by Epstein-Barr virus¹⁰⁹ and beta-human papillomavirus^{110,111}, have also been used to trigger genomic instability. Furthermore, physical stressors like UV irradiation¹¹², ionizing radiation¹¹³, and other external stimuli¹¹⁴ have been used to induce DNA damage, micronuclei formation, and ultimately chromothripsis in diverse cellular systems.

Notably, all these models rely on the artificial induction of chromothripsis. Nevertheless, these models collectively provide researchers with a robust toolkit to study the process of chromothripsis and its implications.

In addition to *in vitro* models, researchers have developed *in vivo* models to gain a deeper understanding of chromothripsis in model organisms^{58,84,96}. Previous research from our lab

Introduction

showed that defective DNA damage repair in mice and human tumours resulted in frequent chromothripsis, linked to *MYC* amplification⁸⁴. Moreover, Simovic et al. used a patient-derived xenograft (PDX) model of medulloblastoma to show that carbon ion radiotherapy can eradicate tumours with chromothripsis^{58,96}.

In the recent years, increased numbers of human studies have been carried out^{23,24,33,51,115,116}. A large-scale analysis of 2,658 human cancers was performed by Cortés-Ciriano et al. using WGS, revealing that chromothripsis is more prevalent than previously estimated, affecting nearly half of the analysed cancers across several cancer entities²³. Different studies have employed a range of sequencing techniques, consistently highlighting the widespread prevalence of chromothripsis in both paediatric and adult tumours, while gaining insight into its role in tumourigenesis and disease progression^{23,24,33,51,115,116}.

Lastly, computational models incorporating artificial intelligence are being developed to identify complex genomic rearrangements, as demonstrated by Yu et al., who applied deep graph learning to detect chromothripsis in multiple myeloma patients¹¹⁷.

1.11. Li-Fraumeni syndrome as a paradigm to study chromothripsis

Given the strong link between *TP53* mutations and chromothripsis, Li-Fraumeni syndrome (LFS) – caused by germline variants in *TP53* – offers a unique model to study chromothripsis^{23,29–33,49,118}. LFS is a rare inherited cancer predisposition disorder, which is characterized by an increased risk of developing a wide variety of cancers at an early age^{118,119}. LFS was first described in 1969 by Li and Fraumeni, who identified families with a history of multiple cancers, which exhibited an inherited pattern of susceptibility^{118,119}. LFS is associated with a broad spectrum of cancers, including soft tissue sarcomas, breast cancer, brain tumours, adrenocortical carcinoma, osteosarcomas, and leukaemia, among others^{118,119}. Patients with LFS have a high lifetime cancer risk, reaching 70% in men and 90% in women¹¹⁸. The onset of these cancers typically occurs at a young age, often in childhood or early adulthood, highlighting the severe impact of defective *TP53* function on early cancer development¹¹⁸.

LFS serves as an ideal model for studying chromothripsis, as the absence of functional p53 in LFS patients allows cells to bypass normal cell cycle checkpoints, which may promote genomic instability^{33,49}. Moreover, LFS-associated tumours frequently exhibit complex genomic rearrangements characteristic of chromothripsis^{33,49}. In line with this, the prevalence of chromothripsis in SHH medulloblastoma with *TP53* mutations in LFS has been reported to be close to 100%³³.

Research using LFS models has provided valuable insights into the mechanisms of chromothripsis. For instance, studies on LFS-primary derived fibroblasts have demonstrated their ability to spontaneously bypass telomere and growth crises in culture, resulting in

chromothripsis^{49,120,121}. Moreover, telomere dysfunction has been shown to play a key role in triggering chromothripsis in such cells⁴⁹.

In addition to *in vitro* models, numerous mouse models and human studies based on LFS have further advanced our understanding of chromothripsis^{33,96}. In conclusion, LFS serves as a paradigm for studying chromothripsis, providing unique insights into its mechanisms, consequences, and potential therapeutic strategies.

2. Aim of the study

Despite extensive research on chromothripsis and genomic instability, the precise factors and triggers rendering cells prone to this phenomenon remain elusive. While several mechanisms have been proposed to explain chromothripsis, key aspects are still unclear. These include the exact timing of events and the molecular basis of this catastrophic event, particularly in its early stages leading to chromosomal fragmentation. Furthermore, the role of telomeres in chromothripsis initiation and their subsequent stabilization in surviving cells, which enables proliferation and evasion of further chromosomal damage, is a critical area of investigation. Thus, in my thesis, I aim to address two central questions:

1. What leads to chromothripsis? In this part, I aim to dissect the underlying mechanism causing chromothripsis.
2. What gives the chromothriptic cells a selective advantage and how do chromothriptic clones become dominant in a cell population?

To answer these questions, I performed a longitudinal analysis to capture the process of chromothripsis *in vitro*. Several *in vitro* models have been previously established, in order to study chromothripsis and its features^{63,64,66,68,74,98,104–114}. However, these models rely on the artificial induction of chromothripsis and may not fully reflect spontaneous chromothripsis. Additionally, early events in chromothripsis-driven cancers are challenging to capture in humans, as most cancer studies have focused on fully developed tumours^{23,24,33,51,115,116}. Therefore, I used a unique model, namely skin-derived fibroblasts from LFS patients, which show spontaneous chromothripsis, eventually becoming immortalized without artificial intervention^{49,122,120}.

In this project, I employed a multi-disciplinary approach to comprehensively investigate the complex phenomenon of chromothripsis. By combining phenotypic profiling, advanced single-cell sequencing techniques, and bioinformatics analyses, I aimed to trace clonal evolution from the pre-neoplastic state up to characterizing the dominant chromothriptic clones. Furthermore, I integrated different layers of multiomics analyses to identify key alterations and mechanisms that lead to this type of genomic instability. Based on the longitudinal profiling, cellular processes were identified, which possibly play a major role in chromothripsis initiation and progression.

Finally, I collaborated with Wisam Zaatra from the Kerem lab to study the mechanisms promoting chromothripsis. We hypothesized that the perturbation of the DNA replication is the underlying mechanism of DNA damage and genomic instability following the loss of WT p53 function.

3. Results

3.1. LFS *in vitro* model

In this project, I used an *in vitro* model of primary skin-derived fibroblasts from LFS patients. These fibroblasts were initially collected from different patients and first reported in a publication from John B. Little's lab at the Harvard School of Public Health in the late 1980s¹²². Later, Michael A. Tainsky's lab (Detroit, USA) extended the culturing, during which the cells underwent a growth crisis¹²⁰. This eventually led to their spontaneous immortalization and chromosomal instability was detected through karyotyping¹²⁰. Cells from four different patients (LFS041, LFS087, LFS172 and LFS174) at different time points were kindly provided by Michael A. Tainsky's lab to our lab. Notably, all patients were in their twenties at the time of sample collection and were not undergoing therapy at the time of sample collection¹²⁰. All four patients had developed cancer twice¹²⁰. Three patients (LFS041, LFS087, and LFS172) had undergone cancer treatment, while LFS174 had not received cytotoxic treatment prior to the skin biopsy¹²⁰.

I used four main time points, enabling a longitudinal analysis of fibroblast behaviour from pre-crisis to immortalization stages. These included early (pre-crisis) cells with typical fibroblast characteristics and normal growth; crisis (during growth crisis), a phase which has been previously characterized by a significant decline in proliferation rates and increased levels of cell death^{1,120,123–126}; post-crisis (after spontaneously bypassing the crisis), where cells exhibit rapid proliferation and altered morphology; and late passages (post-immortalization), defined by clonal dominance long after the chromothriptic event (Figure 7).

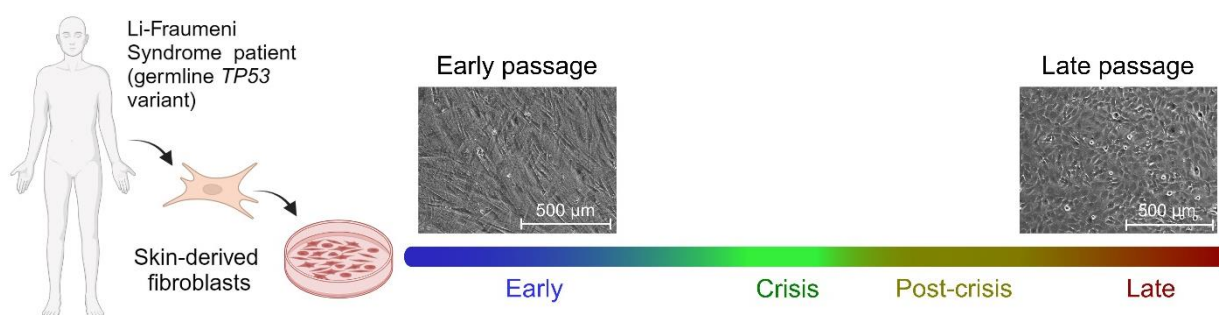


Figure 7. Schematic overview of the model system. The figure describes the *in vitro* model system, derived from skin fibroblasts of LFS patients, which I used to perform a longitudinal study across four time points. Figure adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Original figure created with BioRender.com and Affinity Designer 2.

It is important to note that for early passages of LFS cells, a clonal heterozygous *TP53* mutation is present in one allele, with *p.D184fs* in LFS041^{121,127}, *p.R175H* in LFS172 and LFS174¹²⁷, and *p.R248W* in LFS087^{121,127}, as identified through previous research and by Wisam Zaatra. Our collaboration partners from Batsheva Kerem's lab demonstrated that the function of the

Results

WT p53 is lost during the crisis phase ("second *TP53* hit"). This was confirmed at both the DNA level through sequencing and at the protein level via Western blot analysis (data not shown). These findings indicate that LFS cells undergo a "second hit" to the *TP53* gene, resulting in complete loss of p53 function by the time they reach the crisis stage.

3.2. Morphology and growth characteristics of LFS fibroblasts at different time points

At early passages, fibroblasts from all patients showed normal growth rates, reaching confluency within 7–10 days and were typically split at a ratio of 1:3. Cell growth remained consistent up to passage 25 (p.25) for LFS041, p.26–p.28 for LFS087 and LFS174, and p.40 for LFS172, reflecting differences between the four primary fibroblast cultures as they approached the crisis phase (Figure 8).

During the crisis phase, cell growth slowed down significantly, and most cells entered senescence or apoptosis, which eventually led to widespread cell death. The extensive cell death made it challenging to achieve sufficient cell density, necessitating a lower splitting ratio (1:2) to preserve viability and growth.

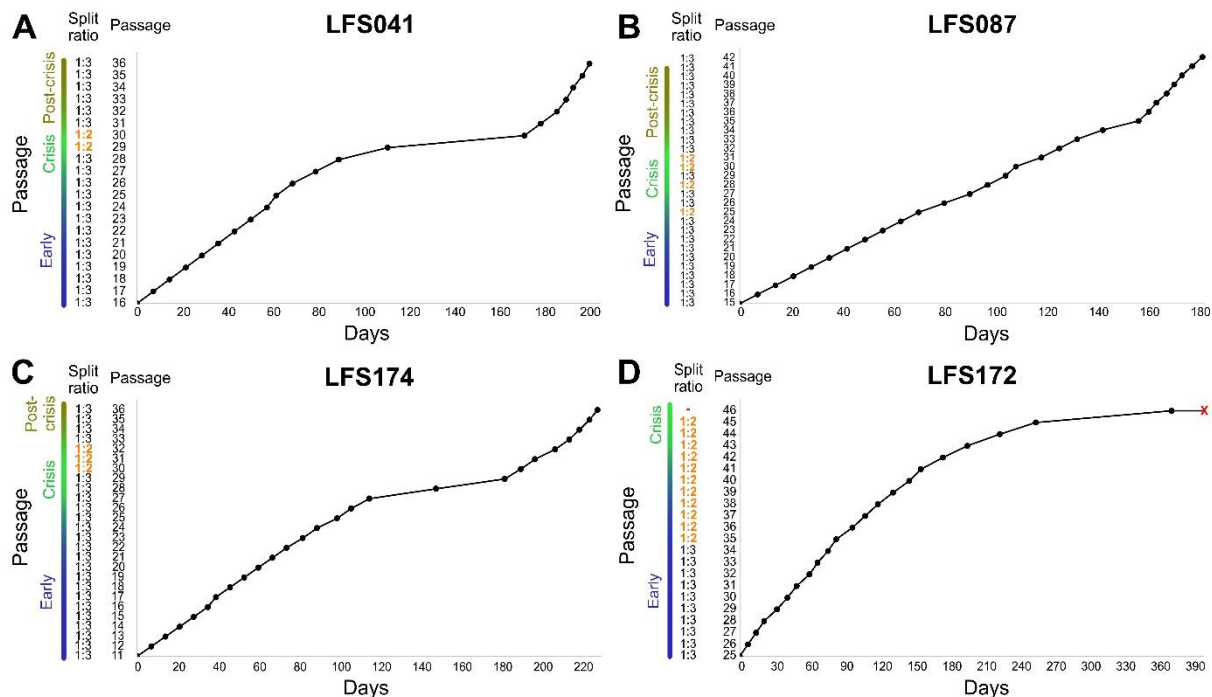


Figure 8. Growth curves of fibroblasts from all four patients. The growth curves illustrate cellular progression from early passage through growth crisis to post-crisis stages for patients LFS041 (A), LFS087 (B) and LFS 174 (C). Cells from patient LFS172 (D) failed to bypass the crisis phase, with the red 'X' marking the point of complete cell death. The x-axis represents time to reach confluency, while the y-axis shows passage number. Split ratios for each passage are indicated on the y-axis, with a splitting ratio of 1:2 highlighted in orange. Panels A and B adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I).

Despite the severe growth arrest during the crisis, one or a few cells may bypass this stage and begin to proliferate again. Cells from LFS041, LFS087, and LFS174 successfully overcame the crisis phase, while LFS172 cells failed to overcome the crisis and stopped growing around passages 45–46. It is important to note that during the crisis, most culture plates experienced extensive cell death. However, the crisis phase typically spans several passages, with some variation between patients. For LFS041, LFS087, and LFS174, passages 26–33 were generally considered crisis passages, whereas for LFS172, the crisis phase occurred later, around passages 41–46. Importantly, even within the same culture, the exact timing of entering the crisis could differ by 1–3 passages from plate to plate, highlighting the heterogeneity not only between patients but also across cells within individual patients.

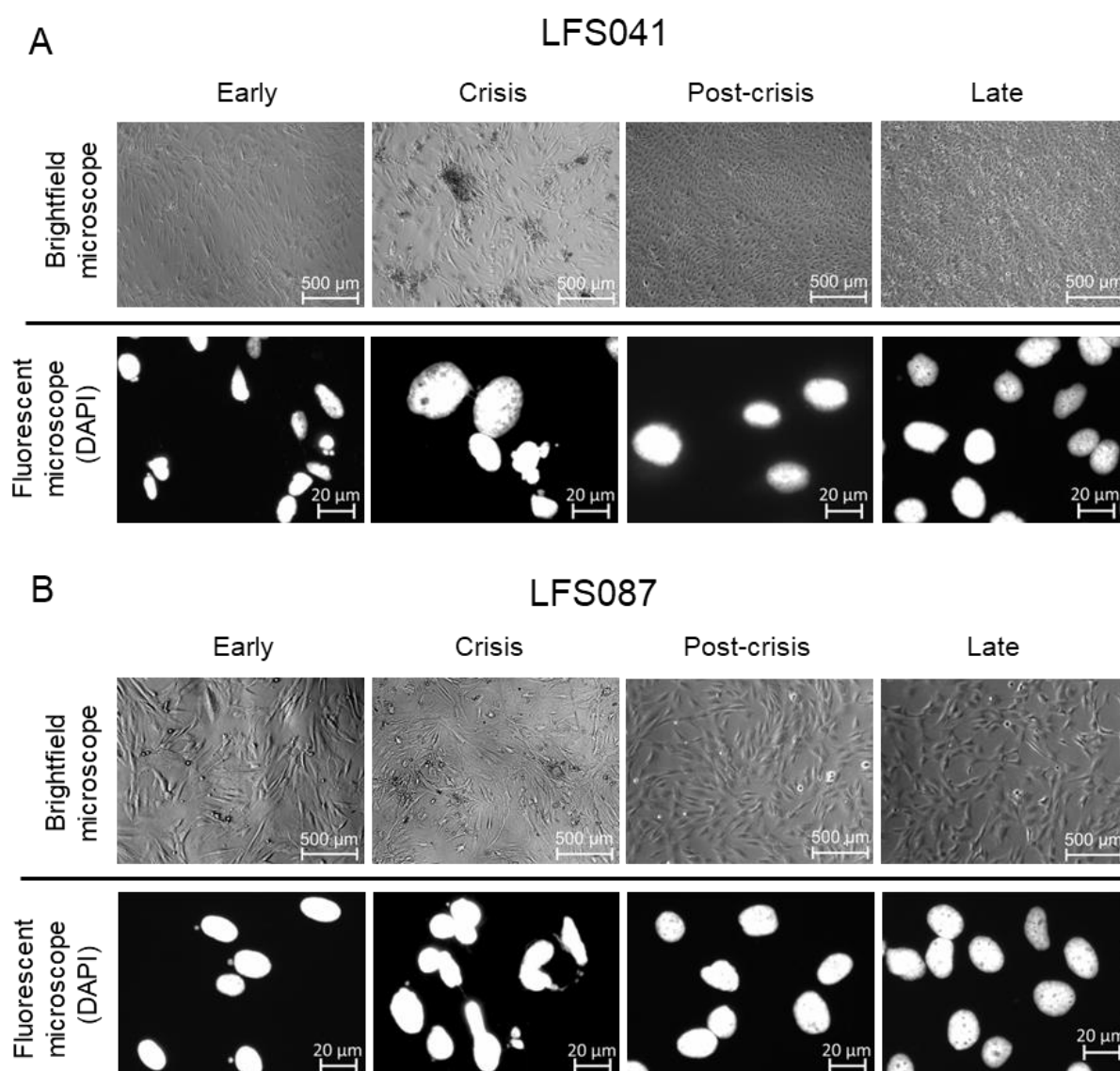


Figure 9. Cellular morphology and DAPI staining of LFS primary fibroblasts across different passages. A, B. Representative images showing cellular morphology (upper panels) and DAPI staining (lower panels) from different passages (early, crisis, post-crisis, and late) of two LFS patients, LFS041 (**A**) and LFS087 (**B**). Brightfield images were acquired using a 5x magnification objective on an inverted microscope (scale bar: 500 μm), while immunofluorescence images were taken using a 63x magnification objective (scale bar: 20 μm). Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I).

Results

Post-crisis (passages 40–80) and late-passage (>80) cells that survived demonstrated rapid expansion. These cells were split at a 1:3 ratio, reaching confluency in just 3-4 days, and could even tolerate higher splitting ratios (e.g. 1:10 ratio), reflecting their enhanced growth rates. In early passages, fibroblasts exhibited a characteristically elongated and flat morphology under a brightfield microscope (Figure 9A-B, upper panels). When stained with DAPI and viewed under a fluorescence microscope, most nuclei appeared small and round, with occasional enlarged cells, chromosomal bridges, and micronuclei. However, during the crisis phase, cell morphology became more heterogeneous. In addition to normal fibroblasts, the cultures contained enlarged, apoptotic and necrotic cells, along with increased cell debris and a subset of smaller, morphologically altered cells. These crisis-associated morphological changes were evident in nuclear morphology, where I observed markedly increased cellular deformation, enlargement, micronuclei formation, and chromatin bridges.

In the post-crisis and late passages, cells displayed a loss of typical fibroblast morphology, with cells becoming more spherical and aggregated. Although these passages showed less nuclear deformation compared to early and crisis passages, nuclei were generally larger. Notably, fibroblasts from each of the four patients displayed distinct morphological alterations, especially in the post-crisis and late passages, reflecting patient-specific features. Altogether, growth curves and microscopy analyses reveal morphological and behavioural changes across passages, which are reflected in the nuclei, indicating increased genomic instability.

3.3. Chromatin bridges and micronuclei as phenotypic characteristics of genomic instability

Driven by observations of nuclei and recent research linking chromatin bridges and BFB cycles to chromothripsis^{49,59–66,128}, I aimed to investigate the frequency and timing of chromatin bridge formation across different passages (Figure 10A). Interestingly, I detected a high number of chromatin bridges even at early passages in patient LFS041 (mean $\pm$ SEM: 29.7% $\pm$ 1.1%). This frequency further increased to 35.6% of the cells ($\pm$ 1.2%) during the crisis (p.27). However, in late-passage cells, the frequency dropped significantly, with less than 2% (SEM: $\pm$ 0.4%) of all examined cells displaying chromatin bridges (Figure 10B). I observed a similar tendency of a gradual increase in LFS087 (Supplementary Figure 1A). Although early-passage cells in LFS087 did not exhibit a frequency of chromatin bridges as high as in LFS041, I noticed a significant increase during crisis passages. Similar to LFS041, post-crisis and late-passage cells in LFS087 showed significantly fewer chromatin bridges (mean $\pm$ SEM: 5.1% $\pm$ 1.3%, 4.5% $\pm$ 1.1%, respectively).

To complement the phenotypic analysis by genomic profiling, I collaborated with Karen Grimes from Jan Korbel's lab to perform Strand-seq and with Petr Smirnov to carry out the

bioinformatics analysis. The experiment was performed on three different time points for patient LFS041: early (p.22, n = 73 cells), post-crisis (p.63, n = 49 cells) and late (p.343, n = 59 cells). In addition to identifying copy number changes, such as deletions and duplications, this sequencing method enabled me to observe different types of SVs and genomic rearrangements with strand-specific information, including inversions, inverted duplications, translocations and complex patterns^{53,54}. In line with these observations, I noticed the presence of fold-back inversions, followed by terminal deletions (Figure 10C upper panel), which is a characteristic of BFB cycles⁵⁵.

To quantify these observations, I calculated the frequency of fold-back inversions across each chromosomal arm in cells from all three sequenced passages. Remarkably, a high proportion of early-passage cells showed fold-back inversions, which were present on different chromosomes at varying frequencies (Figure 10C lower panel).

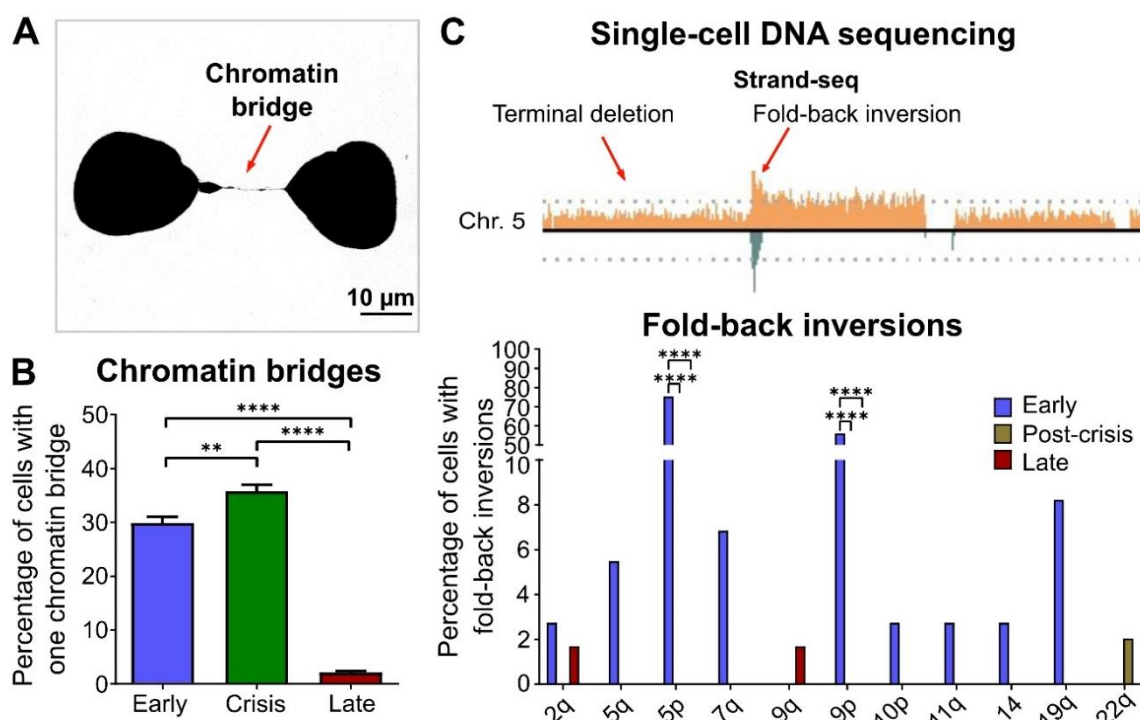


Figure 10. Phenotypic profiling identifies increased frequencies of chromatin bridges and fold-back inversions already at early and crisis passages. **A.** Representative image of a chromatin bridge between two cells. Scale bar: 10 µm. **B.** Quantification of chromatin bridges in LFS041. Four independent biological replicates were performed for each condition, with 750 cells quantified per replicate. Data are presented as mean ± SEM. Statistical significance was assessed using a one-way ANOVA followed by Tukey's multiple comparisons test. **C.** Upper panel: representative example of a fold-back inversion followed by a terminal deletion on chromosome 5 (Strand-seq data). Orange and green represent Watson (forward/plus) and Crick (reverse/minus) strands, respectively. Lower panel: quantification of fold-back inversions in early, post-crisis and late passages from Strand-seq data. Fisher's exact test was performed for pairwise comparisons between each of the three passages for each chromosome arm. P-values were adjusted for multiple comparison tests. For **B** and **C**, p-values lower than 0.05 were considered statistically significant. (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Strand-seq experiment was carried out by Karen Grimes, and bioinformatics analysis was performed by Petr Smirnov.

Results

These differences were significantly higher on chromosomes 5p and 9p in early passages, with 75.3% and 56.2% of cells showing fold-back inversions, respectively, compared to post-crisis and late passages. In contrast, post-crisis and late-passage cells rarely showed fold-back inversions, with frequencies not exceeding 2% of all cells for any chromosome.

In addition to chromatin bridges and fold-back inversions, I also quantified micronuclei, which can serve as a source of complex genomic rearrangements and have been implicated in chromothripsis^{65,66,68,70,71,74}. Interestingly, micronuclei were frequent in early-passage cells of both LFS041 and LFS087 and peaked during the crisis phase. However, the frequency of micronuclei decreased significantly in post-crisis and late-passage cells for both patients, with fewer than 10% of cells forming micronuclei (Figure 11 and Supplementary Figure 1B-C). These observations showed a pattern similar to that of chromatin bridges, with a gradual increase in micronuclei frequency from early to crisis passages, followed by a marked decrease at later passages.

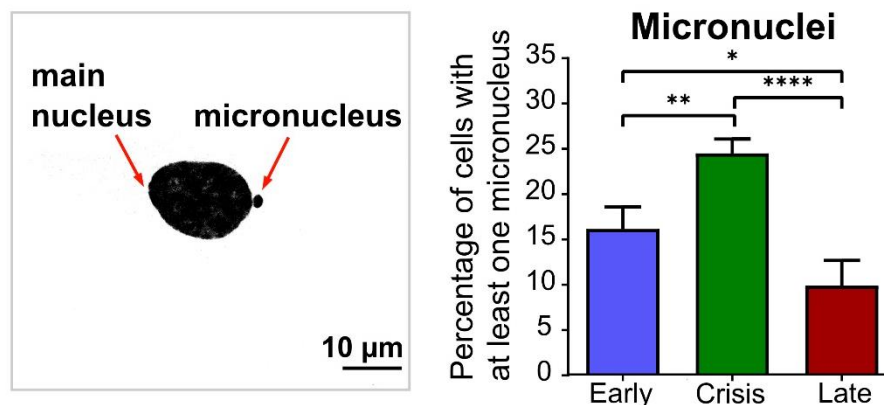


Figure 11. Dynamics of micronuclei formation across passages in LFS cells. Left panel: representative image of a micronucleus. Scale bar: 10 μm. Right panel: quantification of micronuclei in LFS041. Four independent biological replicates were performed for each condition, with 750 cells quantified per replicate. Data are presented as mean ± SEM. Statistical significance was assessed using a one-way ANOVA followed by Tukey's multiple comparisons test, p-values lower than 0.05 were considered statistically significant (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I).

3.4. Characterization of mitotic defects in LFS cells

Recent research demonstrated that mitotic defects contribute to chromothripsis by promoting chromosomal instability through mechanisms such as lagging chromosomes, missegregation, and the formation of micronuclei^{21,29,65,66,129}. Thus, to study the impact of mitotic errors in promoting spontaneous chromothriptic events, I performed double immunostaining of phospho-histone H3 (pH3) and acetyl-α-tubulin.

The analysis revealed various mitotic defects in LFS cells, including failure to congress at the prometaphase, and lagging chromosomes or chromatin bridges (also referred to as anaphase bridges²⁹), and multipolar and monopolar spindle formation (Figure 12). Notably, some cells exhibited extreme patterns with more than one type of defects (Supplementary Figure 1C).

Approximately half of early-passage LFS041 fibroblasts displayed multipolar spindles (mean $\pm$ SEM: 50.3% $\pm$ 4.4%), which significantly increased to 72.3% (SEM: $\pm$ 2.4%) during the crisis, leaving only 22.3% (SEM: $\pm$ 2.6%) of cells in normal mitosis (Figure 12). Multipolar spindle formation decreased markedly in late-passage cells compared to earlier passages. Similarly, in LFS087, crisis-passage cells showed nearly a 30% increase in multipolar spindle formation in comparison to early, post-crisis, and late passages (Supplementary Figure 1D). In summary, the results underscore the link between mitotic defects, particularly multipolar spindle formation, and chromosomal instability, which may contribute to the mechanisms underlying chromothripsis.

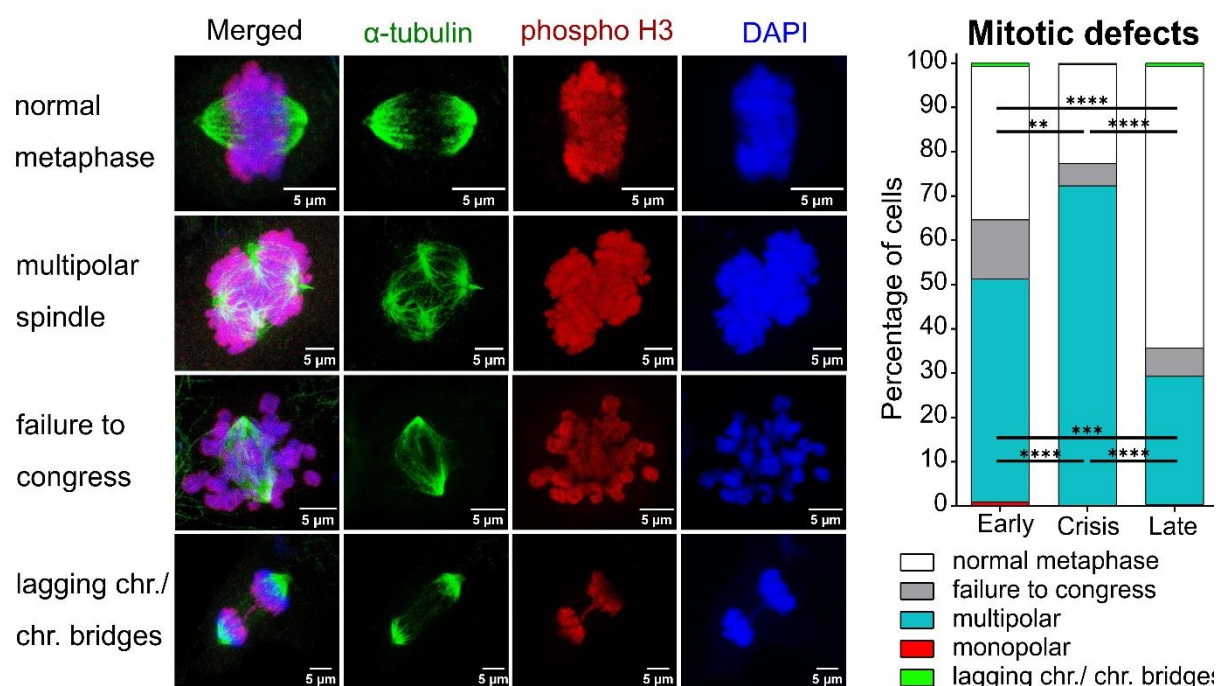


Figure 12. Mitotic defects in LFS cells with increased multipolar spindle formation during crisis. Left panel: representative immunofluorescence images showing a normal metaphase and different types of mitotic defects (multipolar spindle formation, failure to congress and lagging chromosomes or chromatin bridges). Scale bar: 5 μ m. Right panel: quantification of mitotic defects in LFS041. Three independent biological replicates were performed for each condition, with 100 cells quantified per replicate. Data are presented as mean values. Statistical significance was assessed using repeated-measures two-way ANOVA followed by uncorrected Fisher's LSD for multiple comparisons. P-values lower than 0.05 were considered statistically significant. (* p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001). Lagging chr./chr. bridges: lagging chromosomes/ chromatin bridges. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I).

3.5. Comprehensive DNA profiling and SV analysis

3.5.1. Comparative analysis of SV profiles using Strand-seq

Following phenotypic characterization, which was reflected in the altered cellular morphology, I sought to explore the evolutionary dynamics, as well as the spectrum and the timing of genomic alterations occurring across passages. I then used Strand-seq data from the three passages of LFS041 (early, post-crisis, and late) to investigate the timing of SV occurrence

Results

and how clones with different types of alterations evolve over time. For this analysis, Petr Smirnov performed the bioinformatics analysis. I first examined the genomic positions identified through the MosaiCatcher v2 pipeline. To ensure high-confidence results, we applied a stringent filtering criteria, retaining only SVs detected in at least 10% of cells in the resulting heatmap (Figure 13). This threshold minimized the inclusion of noisy and false positive SVs and resulted in a heatmap highlighting SVs with higher confidence.

Interestingly, the analysis revealed a distinct pattern of alterations across time points. Early-passage cells exhibited a predominance of deletions and inversions, particularly on chromosomes 5, 7q and 9p, while chromosomes 3, 6, 11, 15 and 21 showed low numbers of SVs. In contrast, post-crisis and late-passage cells displayed a marked increase in duplications, inverted duplications, and complex structural rearrangements. This shift likely reflects the dynamics of genomic instability and clonal selection of cells overcoming the crisis phase.

Moreover, post-crisis and late-passage cells showed increased clonality, while early passage cells demonstrated greater genomic diversity with a high number of rare events. This pattern became even more evident when applying less stringent filtering criteria, allowing the inclusion of all variants irrespective of their frequency across cells (Supplementary Figure 2). While this approach introduced the possibility of false positives, it allowed the detection of a broader spectrum of SVs, including those present at lower frequencies, potentially capturing events occurring in a single cell. Overall, these results highlight the genomic evolution of LFS041 cells, suggesting a complex and heterogeneous genomic landscape.

3.5.2. Dissecting evolutionary dynamics using SV-based clustering

To gain deeper insights into the dynamics of clonal evolution and SVs over time, I aimed to identify key differences and shared events both across different passages and between individual cells using Strand-seq. This analysis enabled the reconstruction of a detailed phylogeny, providing a framework to trace lineage relationships and uncover the evolutionary trajectory of distinct subclones within the population.

However, during SV analysis using MosaiCatcher v2 on Strand-seq data, I noticed a high rate of apparent false positive calls, inconsistencies, and contradictory SV types in adjacent genomic regions. Thus, to address these limitations and ensure a more accurate analysis, I generated a matrix of five manually called SV types – deletions, inversions, duplications, inverted duplications, and complex rearrangements. For each chromosomal arm, SVs were mapped to their approximate genomic position (in megabase pair (Mbp) resolution) after examining each cell at every passage. SV-based single-cell clustering was performed by Petr Smirnov, based on the manual matrix that I created.

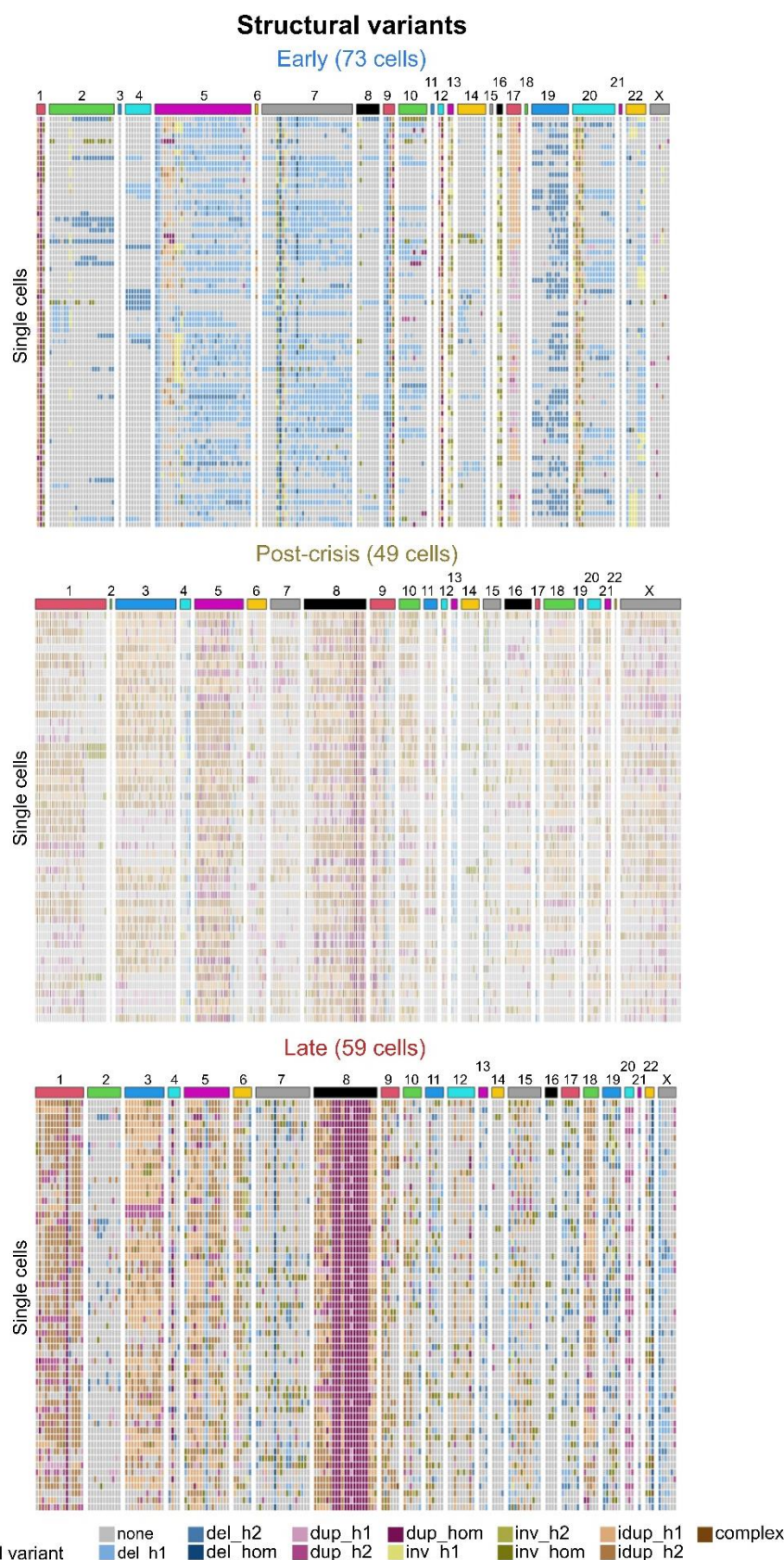


Figure 13. SV profiles across passages in LFS041 using Strand-seq. Heatmaps showing SVs on each chromosome, detected in at least 10% of the cells within each passage: early (p.22), post-crisis (p.63) and late (p.343). Each row shows one cell and each column represents one chromosome. Del: deletion, dup: duplication, inv: inversion, idup: inverted duplication, h1: homolog 1, h2: homolog 2, hom: homologous. Figure adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Petr Smirnov performed the bioinformatics analysis.

Results

We employed two distinct methods to analyse SVs: an event-based approach and a bin-based approach (see methods 6.2.15.3). Showing results from both approaches provides a more comprehensive perspective on the clustering.

3.5.3. Unravelling genomic instability and diversity across cell passages

Trees derived from both approaches (Figure 14A-B) revealed clear clustering of cells by passage, with early-passage cells being the closest to the diploid state, followed by later passages, showing increasingly divergent clustering patterns.

In general, early-passage cells exhibited several distinct subclusters and long branches, indicating greater genetic divergence and variation between cells. Conversely, post-crisis passage cells showed fewer pronounced subclusters and shorter branches, reflecting more genetic homogeneity. Late passage cells, similar to the early passage, displayed long branches and dispersed clusters, suggesting an ongoing evolutionary process and continued genetic divergence over time.

To quantify these observed patterns, we performed two complementary analyses: (1) the calculation of phylogenetic diversity using the total branch lengths within each passage and (2) the assessment of pairwise distance distributions, providing additional insights into the genetic diversity within each passage.

Phylogenetic diversity was estimated by summing the total lengths of the branches in the phylogenetic tree for each passage, thus providing a cumulative measure of genetic divergence within each group. The analysis revealed significant differences in phylogenetic diversity among the three passages. In line with visual observations, the post-crisis passage exhibited the lowest diversity (2791.427), reflecting a relatively homogeneous population with limited genetic divergence. In contrast, early and late passages demonstrated substantially higher genetic divergence, with values of 5071.876 and 5199.626, respectively.

The second analysis, focusing on the distribution of pairwise distances, came in line with the phylogenetic diversity findings (Figure 14C). Post-crisis-passage cells exhibited the narrowest distribution, reflecting a high degree of genetic similarity within this group. In contrast, cells from the early passage showed a broader distribution, with most distances clustering around an intermediate range but also including some outliers with greater genetic divergence. Late-passage cells displayed the widest pairwise distance distribution, corresponding to the highest genetic diversity among the three different time points.

To assess the extent of genomic instability during cellular progression, the number of SVs per single cell based on the SV matrix was also calculated, which allowed me to compare the frequency and variability of SVs across the passages.

The results showed distinct patterns in the number of SVs per cell across passages, with differences in ranges, averages, and variability. In early-passage cells, SVs ranged from 8 to 46 per cell, with a mean of 20 SVs/cell. Notably, this stage exhibited considerable intercellular variability, with some cells showing significant genomic instability and displaying SV numbers comparable to later passages. Post-crisis cells showed an increased SV frequency, ranging from 26 to 50 SVs per cell, with a mean of 38.7 SVs/cell, while late-passage cells demonstrated SV counts ranging from 41 to 73 per cell (mean = 53.2 SVs/cell).

Consistent with SV profiles (Figure 13 and Supplementary Figure 2), these findings demonstrated a progressive increase in SV frequency from early to post-crisis and late passages, in addition to the differences in the observed SV types. Interestingly, the presence of a relatively high number of SVs even in early-passage cells emphasized that cellular stress and genomic instability were already evident at this stage. The wide variability in SV counts during early passages reflected their heterogeneity, with some cells maintained near-normal levels of SVs and others displayed significant instability. In contrast, the higher number of SVs and more clonal SV counts observed in post-crisis and late passages indicate the emergence of dominant clones with enhanced genomic instability.

Collectively, Strand-seq SV heatmaps, phylogeny reconstruction, SV counts per cell, the quantitative analyses of phylogenetic diversity and pairwise distance distributions revealed a decrease in genetic diversity and heterogeneity from early to post-crisis passages, likely driven by selective pressures during the crisis phase. However, the increased diversity observed in late-passage cells suggests a continuous evolutionary process, potentially driven by genomic instability and adaptation during long-term culturing.

3.5.4. Mapping clonal dynamics and their evolutionary trajectories using SVs

To better understand the events driving the evolutionary dynamics and distinguishing the clusters, I manually annotated key differences and shared SVs between the passages and main subclusters. Cells from all passages showed common events, including a small terminal inverted duplication on 12q, alterations on chromosome 5, and a loss on chromosome 9p. However, early-passage cells differed from post-crisis and late-passage cells in several clonal events observed in at least 75% of the cells, including 5p terminal deletion, 5p BFB, 14 telomeric (<10 Mb) deletion and 17q terminal gain. In contrast, post-crisis and late passages shared clonal events absent in early-passage cells, such as polyploidy, 4q deletion, 5p gain with a telomeric deletion, 7p gain, 8q gain, 10q gain and complex patterns on chromosome 15.

Results

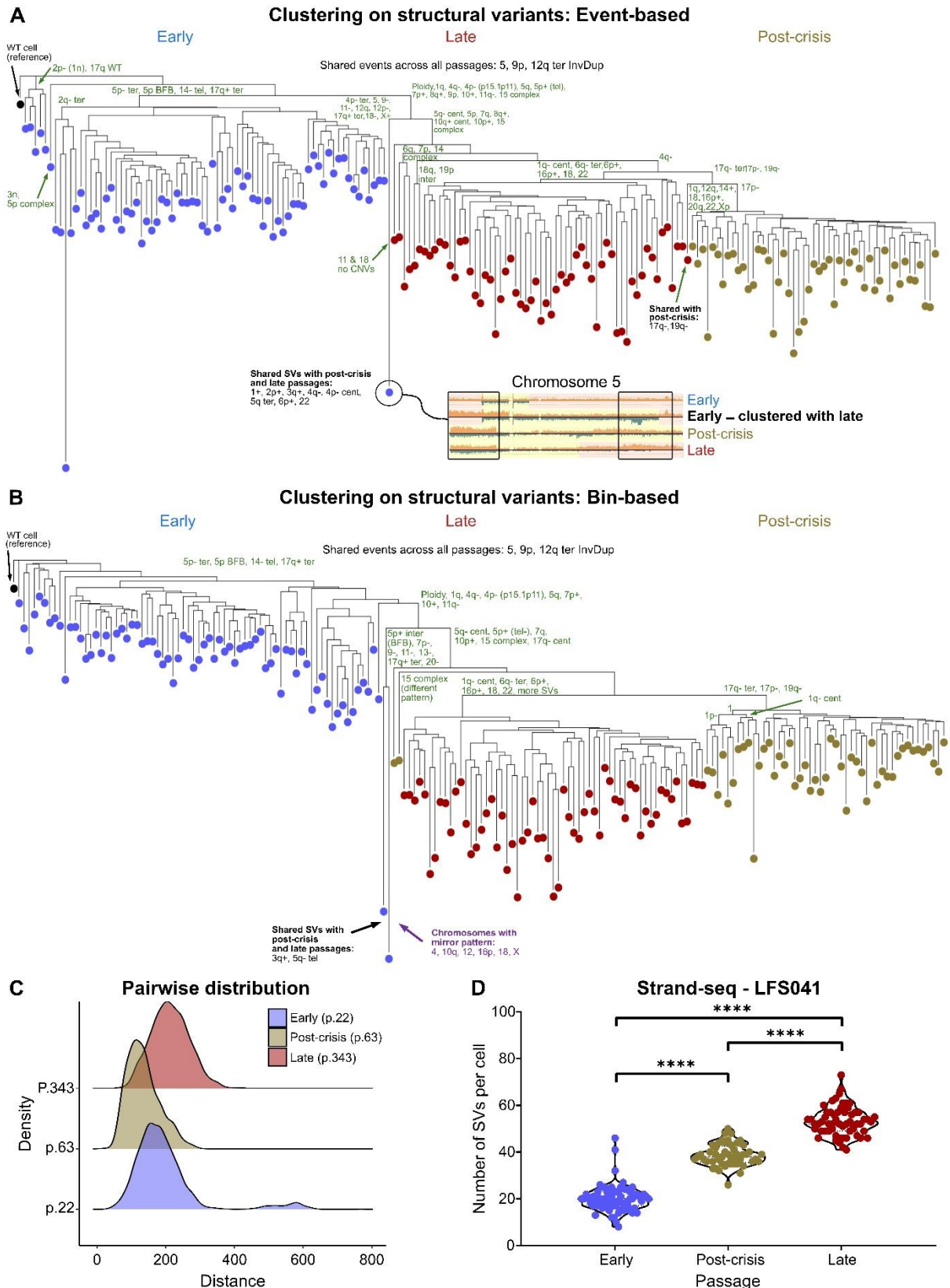


Figure 14. Clustering on SVs and pairwise distance distributions in Strand-seq data. **A.** shows the clustering using an event-based modified Hamming distance and neighbour-joining, while **B.** shows the clustering on SVs detected in Strand-seq data using modified Hamming distances on 2MB bins and neighbour joining hierarchical clustering (Methods 6.2.15.3). SVs separating the main cell clusters are shown on the nodes. Shared events are annotated in black, while differences between the main clusters are annotated in green. Chromosomes with mirror patterns between cells are described in purple. In **A**

the bottom panel exhibits chromosome 5 from representative early-passage, post-crisis and late passage cells. Abbreviations: -: loss, +: gain, cent: centromeric, tel: telomeric, term: terminal. The distance between cells reflects the number of events. **C.** Density plot illustrating the distribution of pairwise distances among fibroblast cells from early (p.22), post-crisis (p.63), and late (p.343) passages. The x-axis represents pairwise genetic distances, while the y-axis indicates the density of cells exhibiting a given distance value. **D.** Number of SVs detected per cell across the three passages: early (n = 73), post-crisis (n = 49), and late (n = 59). Statistical significance was assessed using a nonparametric Kruskal-Wallis test, followed by Dunn's multiple comparisons test. P-values < 0.05 were considered statistically significant (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). Figure adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Petr Smirnov performed the bioinformatics analyses.

Despite these shared features across passages, post-crisis and late-passage cells exhibited notable clonal differences. For instance, 17q terminal loss, 17p loss, and 19q loss were present clonally in post-crisis but not in the late passage. Conversely, late-passage cells displayed unique events such as 1q centromeric loss, 6p gain, and 16p gain, along with an increased number of non-clonal SVs across various chromosomes.

Interestingly, in both clustering methods, I detected one early-passage cell that clustered with post-crisis and late-passage cells, rather than with other early-passage cells (Figure 14A-B). Although this cell shared some SVs with early-passage cells, it lacked clonal key events observed in the early passage, including 5p terminal loss, 5q deletion, 19q deletion, and 22q centromeric deletion. Of note, this cell showed additional differences compared to the cells from later passages, affecting chromosomes such as 4, 5, 9, 11, 15, 17, and X. However, it exhibited similar SVs to those seen in post-crisis and late-passage cells, including gains on chromosomes 1q, 2p, and 3q, 6p gain, 4q loss, 4p centromeric loss, along with a comparable pattern on chromosome 22. Additionally, the cell lacked the 5p terminal deletion and 5q loss seen in other early-passage cells, but exhibited a shared gain on the q arm, which was also observed in post-crisis and late-passage cells (Figure 14A, bottom panel).

Furthermore, the bin-based approach identified one additional early-passage cell clustering with cells from post-crisis and late passages. These two cells together showed limited shared events with the later passages, including only 3q gain and 5q telomeric loss. However, they exhibited a mirror pattern involving whole chromosomes or entire chromosomal arms, specifically across chromosomes 4, 10q, 12, 16p, 18, and X, where a loss in one cell corresponded to a gain in the other.

Altogether, these results highlight distinct structural alterations between and within the different passages, reflecting temporal evolutionary trajectories and suggesting that only a small fraction (1–3%) of early-passage cells display rearrangements typically observed as clonal events in later passages.

Results

3.5.5. Chromothripsis and complex rearrangements at the single-cell level

In a previous study by Ernst et al. (2016), bulk WGS of early passages (p13–19) from all four patients revealed only a few CNAs, with no evidence of chromothripsis at this stage (data not shown)⁴⁹. Conversely, chromothripsis was observed in the late passages (p84–308)⁴⁹. However, bulk sequencing data might mask subclonal or rare events. Therefore, I examined Strand-seq data on LFS041 cells to investigate complex rearrangements such as chromothriptic events, as well as their frequencies across the passages.

At the single-cell level, I detected clonal chromothripsis on three different chromosomes (5, 11, and 15) in post-crisis and late-passage cells (Figure 15A). Additionally, complex rearrangements were identified in 10 post-crisis cells and 5 late-passage cells, affecting other chromosomes. Surprisingly, nearly one third (24/73) of early-passage cells exhibited one or more chromosomes with complex rearrangements and/or chromothriptic patterns, with 6 of these cells showing such patterns on two different chromosomes. These complex rearrangements were distributed across nearly half of the genome (chromosomes 2, 3, 5, 7, 8, 10, 13, 14, 19, 20, and X) at varying frequencies. Particularly, chromosomes 5 and 7 were most frequently affected, with complex rearrangements observed in 10 and 5 early-passage cells, respectively. Two early-passage cells exhibited similar complex patterns on chromosome 5p (Figure 15A, highlighted cells with a single asterisk), potentially resulting from telomere attrition and BFB cycles. Moreover, two other cells exhibited complex rearrangements on 5q, in regions that overlapped with, but were not identical to, those seen in post-crisis and late-passage cells (Figure 15A, highlighted cells with two asterisks).

I next sought to determine, which of these complex patterns seen in early passage persisted as clonal events in post-crisis and late passages and were supported by bulk WGS analysis of post-crisis (p.63) cells. Among the complex rearrangements detected in early-passage cells, only two events showed similar – but not identical – SV patterns to those observed as clonal chromothriptic events in later passages. Specifically, one early-passage cell exhibited complex rearrangements on chromosome 5, and another on chromosome 8, showing patterns similar to those observed in post-crisis and late-passage cells (Figure 15B-C and Supplementary Figure 3).

In summary, single-cell Strand-seq analysis revealed genomic instability, including complex rearrangements in a subset of early-passage cells that were undetectable by bulk WGS. Importantly, most complex rearrangements observed in early passages were not maintained in later passages, suggesting that these cells did not survive or lacked a selective advantage.

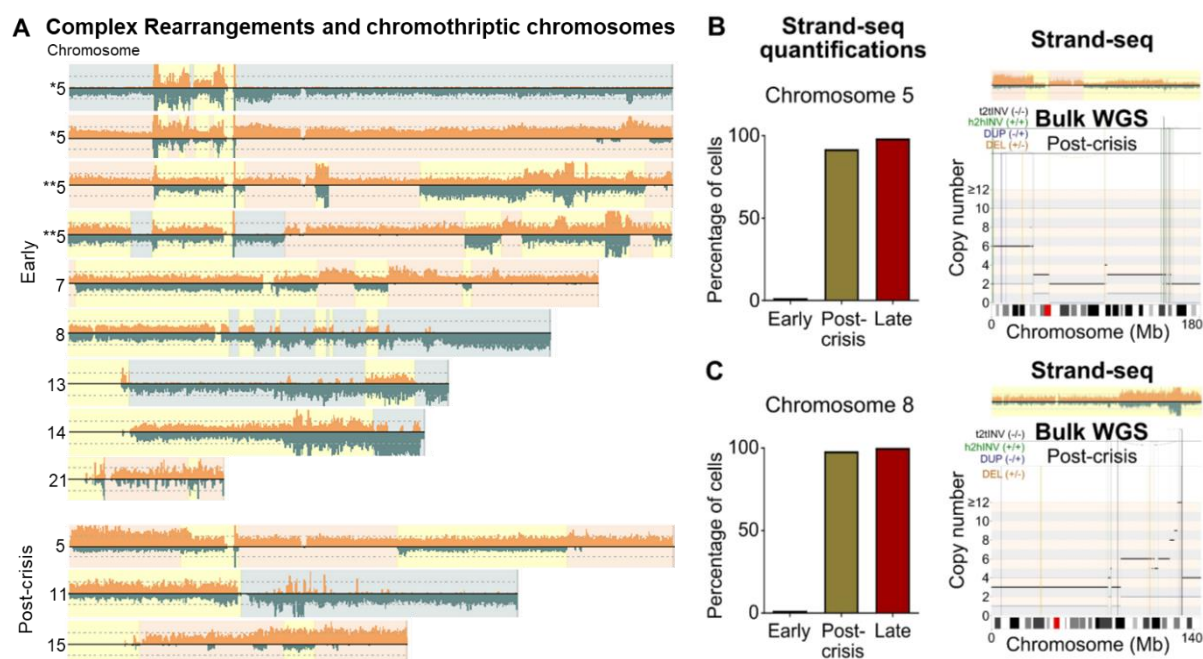


Figure 15. Single-cell Strand-seq and bulk WGS analysis of complex rearrangements in LFS041 cells. **A.** Examples of complex rearrangements and chromothriptic chromosomes from early and post-crisis-passage cells of patient LFS041. The top two chromosomes (marked with one asterisk) exhibit similar complex patterns on 5p between two distinct early-passage cells. The next two chromosomes (marked with two asterisks) show SVs on 5q in two early-passage cells, which overlap with those observed in post-crisis and late-passage cells. **B, C.** Left panels: Quantification of the frequency of similar complex rearrangements on chromosomes 5 (**B**) and 8 (**C**) across different passages from LFS041 based on Strand-seq data. Right panels: Representative examples of complex rearrangements in the Strand-seq data and ReConPlots from bulk WGS (LFS041 p.63, post-crisis passage) for chromosomes 5 and 8. Total chromosome length is shown in megabase (Mb). Abbreviations and colour coding: t2tINV: tail-to-tail inversion (black), h2hINV: head-to-head inversion (green), DUP: duplication (blue), DEL: deletion (orange). Figure adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Petr Smirnov performed Strand-seq bioinformatics analysis and generated ReConPlots from bulk WGS.

3.5.6. Characterizing ploidy states in Strand-seq and metaphase spreads

Several studies have linked genome doubling and chromothripsis in cancer^{33,107,130}. However, the precise temporal relationship and causal links between chromothripsis and whole-genome doubling remain unresolved. Strand-seq data from LFS041 showed that all cells in post-crisis and late passages in LFS041 were polyploid. This pattern was also evident in the pseudobulk analysis, where almost all chromosomes of post-crisis-passage cells were triploid, and chromosomes in late-passage cells had triploid or tetraploid states (Supplementary Figure 4). While the pseudobulk analysis indicated that early-passage cells were diploid, the single-cell level analysis revealed one triploid cell out of a total of 73 cells.

To further validate these findings, I prepared metaphase spreads, and chromosome staining and counting were performed by Frauke Devens at several time points for patients LFS041 and LFS087 (Figure 16).

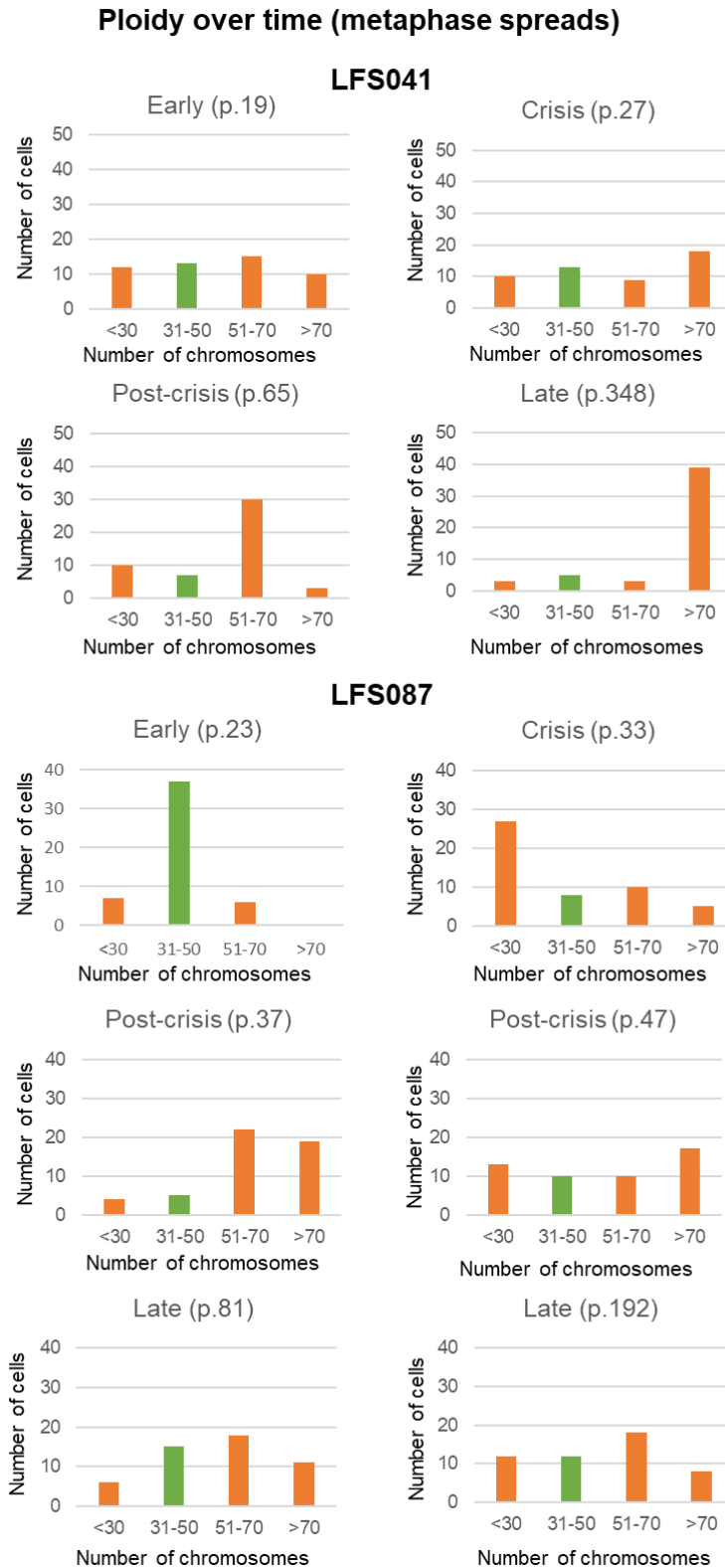


Figure 16. Ploidy assessment of metaphase spreads for patients LFS041 and LFS087 shows polyploidization in post-crisis and late passages as compared to early passage cells. Chromosomal counts were obtained from 50 cells per passage for patients LFS041 (A) and LFS087 (B). Bar graphs display the distribution of chromosome numbers per cell. Green bars represent near-diploid cells (31-50 chromosomes), while orange bars indicate aneuploid populations (<30, 51-70, and >70 chromosomes). Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Chromosome staining with DAPI and ploidy assessment was performed by Frauke Devens.

Ploidy evaluation revealed a high degree of variability in LFS041 cells during early and crisis passages, with cells distributed across haploid, diploid, triploid, and tetraploid states. These ploidy levels were approximately evenly represented, highlighting the heterogeneous nature of the cells and coming in line with the results shown earlier. This variability likely explains the diploid state of early-passage cells observed in the pseudobulk Strand-seq analysis. The averaging effect of combining cells with diverse ploidy states results in a predominant diploid signature, masking the heterogeneity of the population. Consistent with Strand-seq results, cells in post-crisis and late passages showed a predominance of triploid and tetraploid states, respectively. In addition, bulk WGS data from LFS041 p.63 (post-crisis) showed an average population ploidy of 3.65, confirming findings discussed above.

In contrast, for patient LFS087, early-passage cells had a majority of diploid cells (74%). The crisis-passage cells exhibited increased chromosomal loss, with over half of the cells showing a haploid appearance. Remarkably, just four passages later, more than 80% of post-crisis cells were polyploid, with near-equal distribution between triploid and tetraploid states. Although later passages did not show an increase in the frequency of polyploid cells, they still comprised at least 52% of the population in both post-crisis and late passages.

Altogether, despite the differences between LFS041 and LFS087 in early and crisis passages, both cultures show a quick shift towards polyploidization in post-crisis and late passages. Thus, this rapid transition in ploidy status suggests that genome doubling may be a key adaptive response to genomic instability.

3.5.7. High-throughput sequencing reveals key CNAs

Although Strand-seq is a powerful method to detect complex SVs at the single-cell level with strand-specific resolution, it exhibits some limitations, such as low-throughput and unsuitability for high copy number inference. To address these limitations, I utilized HIPSD-seq to analyse CNAs at a high-throughput, which we developed in our lab in collaboration with Jan-Philipp Mallm (W192) and Oliver Stegle (B260)⁵². For the experimental work, I collaborated with Pooja Sant, while Jan Otoničar led the computational analysis.

We applied HIPSD-seq (scDNA) to three different time points of the LFS041 fibroblasts: early (p.22), crisis (p.27), and post-crisis (p.62), as well as to the late passage of LFS087 (p.196). This approach enabled the sequencing of thousands of cells, allowing the detection of rare events, particularly in the early and crisis passages.

UMAP clustering of LFS041 fibroblasts revealed a clear separation between early/crisis passages and post-crisis cells. Notably, early and crisis passage cells predominantly clustered together (Figure 17). However, some cells exhibited a transient stage with overlap between all three passages. Additionally, a small number of cells from early and crisis passage co-

Results

clustered with post-crisis cells. While low coverage and noise may affect these observations, the overlap could also suggest shared CNAs between the crisis and post-crisis passages.

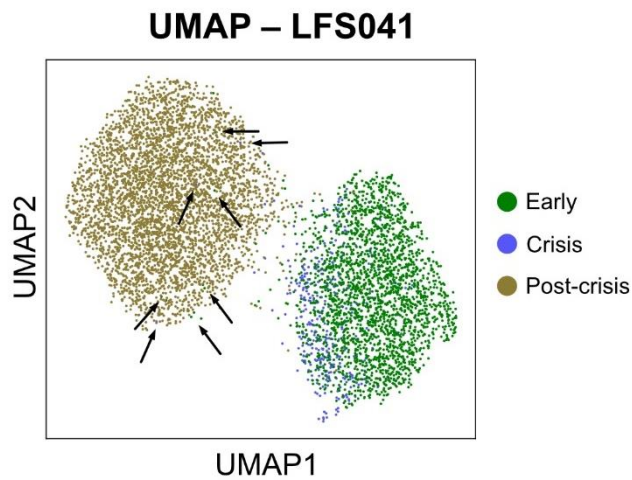


Figure 17. UMAP visualization from HIPSD-seq (scDNA seq) showing clustering of cells across different passages of LFS041. Early (p.22) and crisis (p.27) passages primarily co-cluster, while post-crisis cells (p.62) form a distinct cluster. Black arrows point to a small number of early and crisis-passage cells that are clustered with the post-crisis cells. Jan Otoničar performed the bioinformatics analysis.

For LFS041, copy number plots revealed high heterogeneity and high number of CNAs already present in the early-passage cells, with an even greater number observed in crisis-passage cells (Figure 18). Despite this heterogeneity, clonal events were identified, including 5p loss, with many cells also exhibiting whole-chromosome 5 loss. Additional recurrent alterations included 17q gain and losses of 7q, 9p, and 19q. These regions contain genes that might play an important role in the mechanism of chromothripsis initiation and/or facilitating the survival of chromothriptic cells through the crisis phase, such as the loss of *CDKN2A* locus.

In contrast, post-crisis passage cells showed more clonal CNAs, particularly copy number gains, and less heterogeneity compared to earlier passages. Amplifications were observed at important loci, including the *MYC* locus on chromosome 8q and the *TERT* locus on 5p, both of which were in line with Strand-seq results, with the *TERT* locus amplified in post-crisis passage cells but lost in early and crisis passages.

In summary, the results align with the findings from Strand-seq regarding cellular diversity, clonality, and key events in LFS041. Both early and crisis passages exhibited heterogeneity, whereas post-crisis passage cells demonstrated a shift towards greater homogeneity. Furthermore, HIPSD-seq facilitated the identification of essential focal gains and losses, providing insights into the genomic alterations driving the transition from early to crisis and post-crisis stages.

For LFS087, late-passage cells exhibit clonal gains and losses across different chromosomes in all cells, with a dominant clone comprising approximately two thirds of the overall population. In addition to the dominant clone, I identified two other subclones, each harbouring distinct CNAs compared to the dominant clone (Supplementary Figure 4). Key differences between these three clones were observed on chromosomes 2, 3p, 5, 7, 8, 9q, 10, 11, 13, 15, 16, and

22. Similar to the late passage of LFS041, these findings reflect an ongoing evolutionary process that may be significantly influenced by prolonged cell culturing.

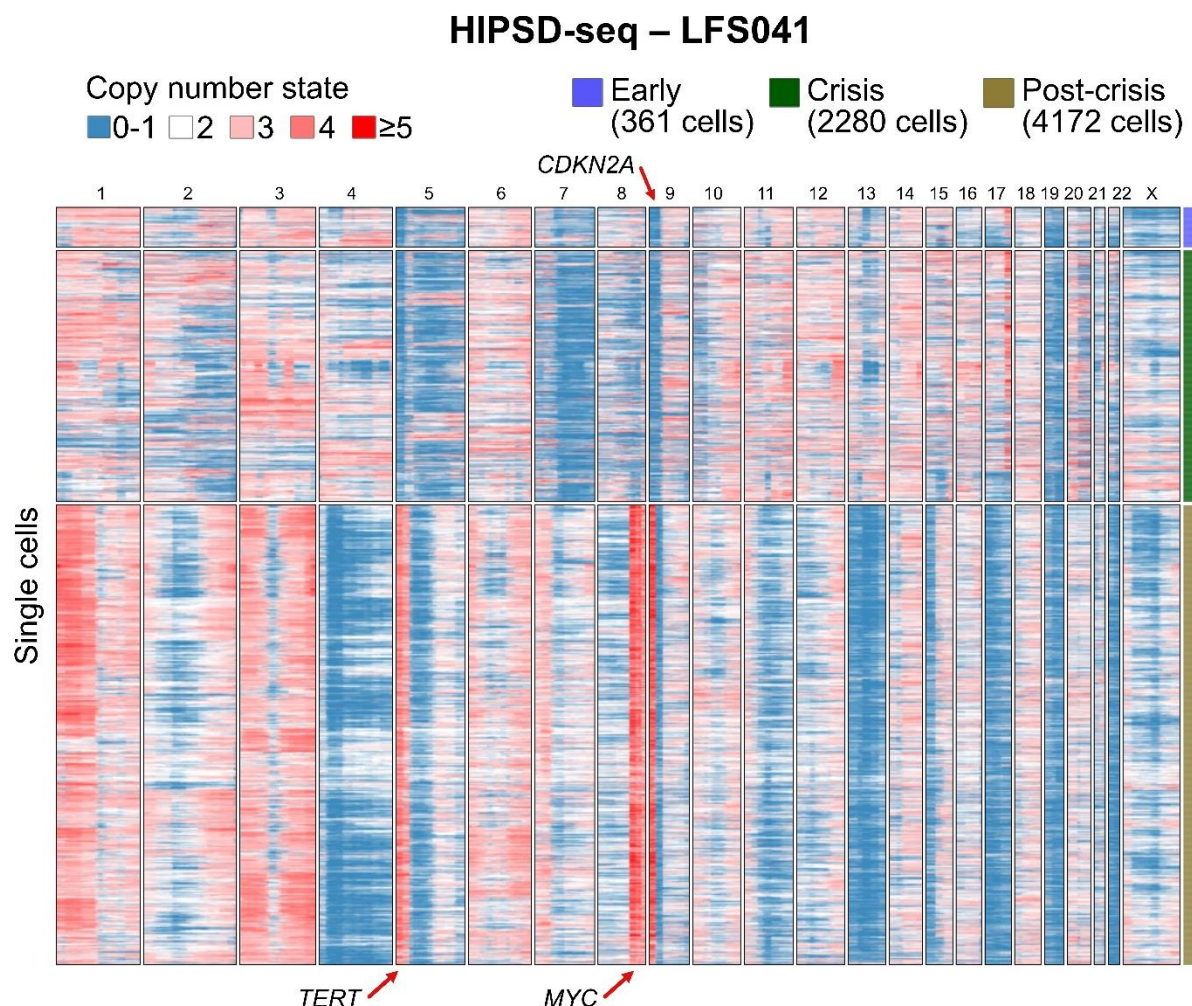


Figure 18. Heatmap representation of copy number profiles in LFS041 samples analysed using HIPSD-seq. Three passages are represented on the heatmap: early (p.22, n = 361 cells), crisis (p.27, n = 2280 cells) and post-crisis (p.63, n = 4172 cells), with a total of 6813 cells. Copy number profiles of post-crisis cells were corrected for ploidy. Each row represents one cell, and each column represents one chromosome with a bin size of 1000 kb. Figure adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Pooja Sant performed the experimental part, and Jan Otoničar performed the bioinformatics analysis.

3.5.8. Bulk WGS analysis identifies chromothriptic chromosomes and key drivers in LFS

The single-cell sequencing analyses that I performed provided valuable insights into the genomic landscape at single-cell resolution. However, due to the low genome coverage and the potential for noise and false-positive calls associated with scDNA sequencing, I performed bulk WGS to validate the single-cell findings and detect chromothripsis with high confidence.

Since the post-crisis passage of LFS041 (p.63) and the late passage of LFS087 (p.196) were predominantly clonal, as previously demonstrated by scDNA sequencing, I performed bulk WGS at high coverage (30-50X) on similar passages; LFS041 (p.63) and LFS087 (p.195). These passages were chosen because early and crisis passages are highly heterogeneous,

Results

and bulk WGS could mask the effects of subclonal populations and rare genomic events. The bioinformatics analysis of bulk WGS was carried out by Petr Smirnov.

ShatterSeq results confirmed the presence of chromothripsis on multiple chromosomes in both LFS041 and LFS087 (Figures 19 and 20, respectively). In LFS041 p.63, chromothripsis was identified on chromosomes 5, 11, 15, 17, and X. Analysis using ShatterSeq and OncoAnalyser revealed patterns of complex genomic rearrangements and alterations, with both inter- and intra-chromosomal rearrangements involving chromosomes 5, 7, 8, 11, 12, 15, and X (Supplementary Figure 6). For LFS087 p.195, chromothripsis was detected on chromosomes 1, 12, 15 and 20, with inter-chromosomal rearrangements involving chromosomes 1, 7, 10, 11, 15, and 20 (Supplementary Figure 7).

Bulk WGS further enabled the identification of breakpoints and genes affected by the rearrangements. Breakpoint analysis revealed that 653 genes were directly affected by breakpoints in LFS041 post-crisis, while 727 genes were affected in LFS087 late passages. Interestingly, 58 genes were shared between the two patients. These included oncogenes such as *TOP2A*, *SLC3A2*, and *SHANK2*, as well as tumour suppressor genes like *INPP5D*, *RBFOX1*, and *NLRP6*, highlighting genes potentially involved in the dominance of the surviving clones or in initiating chromothripsis.

As expected, in LFS041 *TP53 p.D184fs* was identified as a driver, which corresponds to the germline variant in this patient^{121,127}. The average *TP53* copy number was found to be 2, nearly half of the average ploidy of the population (3.65), indicating the loss of the second allele at this passage. Another driver identified was *GRIN2A p.P23Q*, which has been described as a tumour suppressor and has been implicated in various cancers due to its role in glutamate signalling pathways^{131–134}. The analysis revealed *MYC* as a somatic copy number driver, with an average amplification to 11.8 copies per cell. Additionally, *TERT* exhibited high amplification at 10.3 copies. Both of these amplifications were previously identified through single-cell sequencing.

Similarly, in LFS087 p.195, the germline variant *TP53 p.R248W* was identified as a driver. Additionally, *ATRX p.V1002fs* was detected as a second driver. *ATRX* plays a crucial role in telomere maintenance through ALT mechanism⁶². Collectively, through bulk WGS data, I identified several chromothriptic chromosomes in both LFS041 and LFS087, with chromosome 15 being chromothriptic in both patients and nearly 60 shared genes affected by breakpoints.

Bulk WGS – LFS041 p.63 (post-crisis) – ReConPlot (copy number plots)

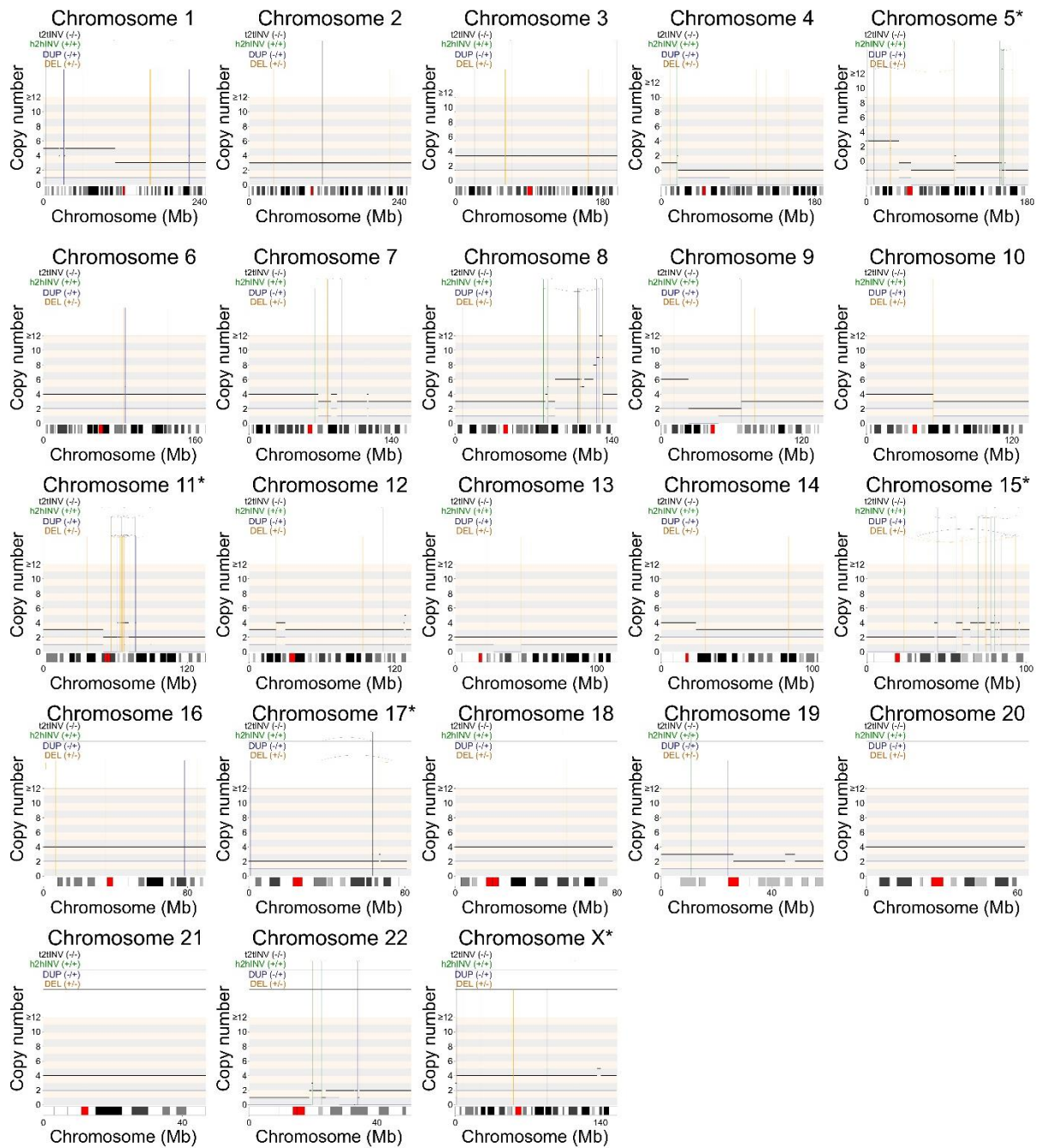


Figure 19. ReConPlots of SVs and somatic CNAs identified from WGS data for patient LFS041 p.63 (post-crisis). Each plot displays the distribution of CNAs along each chromosome, with colour coding indicating sites of deletions (orange), duplications (blue), head-to-head inversions (green) and tail-to-tail inversions (black). Total chromosome length is shown in megabase (Mb). Chromosomes annotated with an asterisk (*) indicate chromothriptic chromosomes. t2tINV: tail-to-tail inversion, h2hINV: head-to-head inversion, DUP: duplication and DEL: deletion. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). ShatterSeek analysis was performed by Petr Smirnov.

Results

Bulk WGS – LFS087 p.195 (late) – ReConPlot (copy number plots)



Figure 20. ReConPlots of SVs and somatic CNAs identified from WGS data for patient LFS087 p.195 (late). Each plot displays the distribution of CNAs along each chromosome, with colour coding indicating sites of deletions (orange), duplications (blue), head-to-head inversions (green) and tail-to-tail inversions (black). Total chromosome length is shown in megabase (Mb). Chromosomes annotated with an asterisk (*) indicate chromothriptic chromosomes. t2tINV: tail-to-tail inversion, h2hINV: head-to-head inversion, DUP: duplication and DEL: deletion. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). ShatterSeek analysis was performed by Petr Smirnov.

3.6. Telomere stabilization in LFS fibroblasts

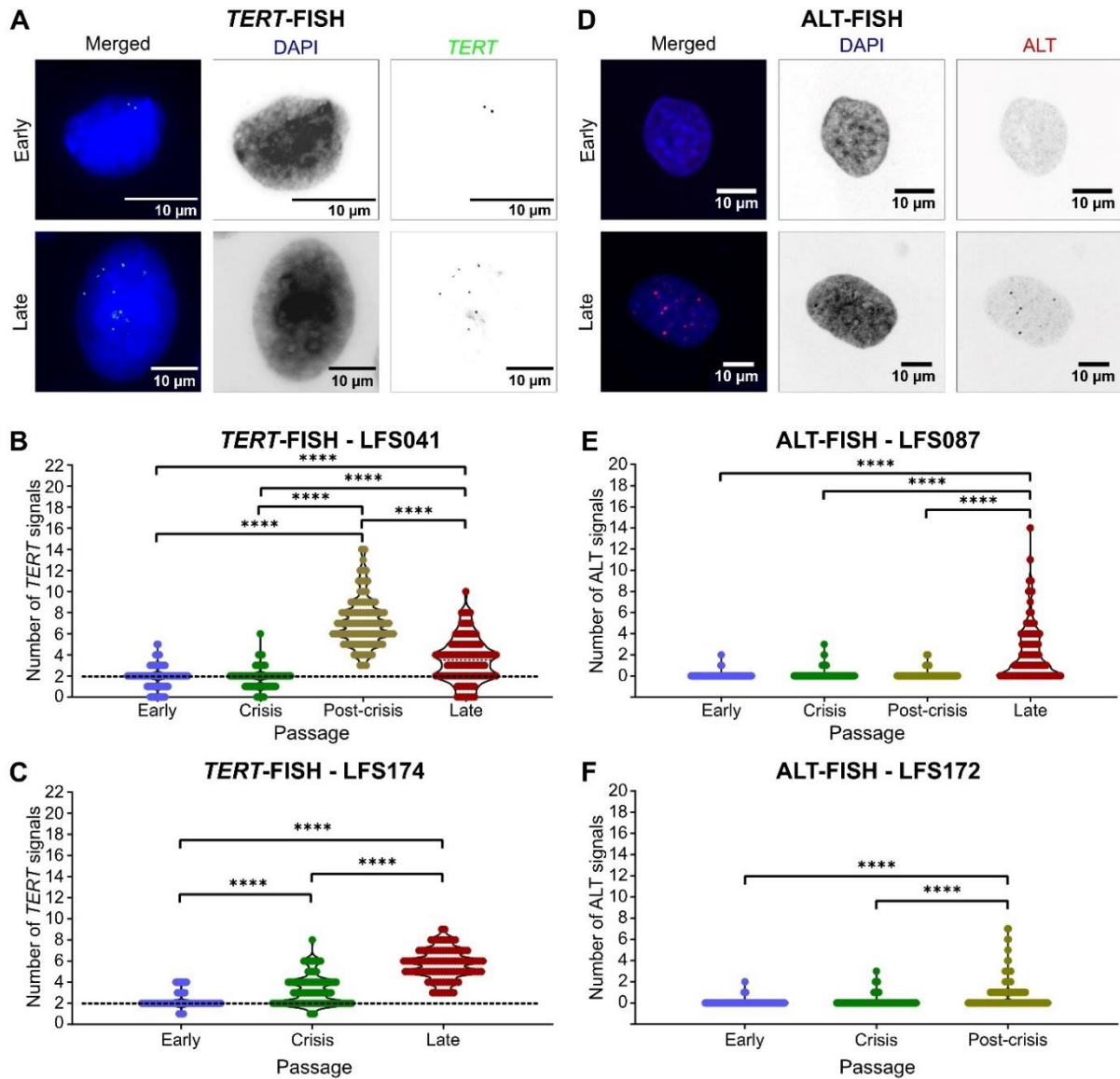
After identifying *TERT* amplification in LFS041 through both single-cell and bulk WGS, as well as a mutation in the *ATRX* gene in LFS087, I investigated the role of telomere stabilization in chromothriptic cells in greater detail. Previous studies have shown that telomere stabilization occurs via *TERT* activation in LFS041 and LFS174, whereas LFS087 and LFS172 achieve telomere stabilization through the ALT mechanism^{49,121,135}. However, earlier work was limited to only two time points per patient, namely early and post-crisis/late⁴⁹.

To pinpoint the exact timing of telomere stabilization and understand how this process contributes to the survival and clonal dominance of chromothriptic cells, two types of FISH experiments were performed. Both *TERT*-FISH and ALT-FISH were performed by Frauke Devens, with Anne-Sophie Alexio contributing to the ALT-FISH experiments. These experiments allowed studying telomere stabilization mechanisms longitudinally, while offering a single-cell resolution to investigate critical time points (Figure 21A-F).

Consistent with previous findings, I detected a significant increase in *TERT* signals in LFS041 post-crisis-passage cells, ranging between 3 and 14 signals per cell (Figure 21B). Late-passage cells showed high heterogeneity, with *TERT* copy numbers varying from complete loss to gains of up to 10 copies. Both post-crisis and late passages demonstrated a significant increase in *TERT* signals compared to early and crisis passages. In contrast, the majority of early- and crisis-passage cells displayed either copy-neutral counts or *TERT* loss. However, even at these earlier stages, some cells harboured gains of *TERT*. LFS174 cells showed a pattern similar to LFS041. While most early-passage cells exhibited two copies of the *TERT* locus, some cells already displayed gains or losses of *TERT*. The increase became significantly more apparent during the crisis passage, where over half of the cells (58%) showed elevated *TERT* copy numbers, with some reaching levels comparable to late-passage cells (up to 8 copies, Figure 21B). In late passages, no cells retained normal *TERT* copy numbers and all cells exhibited amplifications, resembling the post-crisis passage of LFS041.

The majority of early and crisis-passage cells in both LFS087 and LFS172 did not show ALT signals, indicating the absence of ALT activation in most cells. However, a small subset of cells in each time point for both patients displayed 1–3 ALT signals per cell, suggesting localized activation of the ALT mechanism in these cells. Although in LFS087 ALT signal levels in the post-crisis passage were comparable to early passages, this was not the case in the late passage. Nearly two thirds of the cells exhibited a significant increase in ALT signals, with some cells showing more than 10 signals per cell. Similarly, in LFS172, ALT levels were significantly elevated in the late passage compared to early and post-crisis passages.

Results



Despite the significant increase in ALT signals observed in post-crisis passages (LFS172) and late passages (LFS087), a substantial fraction of cells did not show any ALT activity. Specifically, over half of the cells in LFS172 and one third of the cells in LFS087 lacked ALT signals, raising questions about the mechanisms of telomere stabilization in these cells.

To investigate whether other telomere stabilization mechanisms might be involved, I examined the presence of *TERT* signals in the post-crisis passage of LFS172 (p.72). Surprisingly, over 40% of the cells displayed increased *TERT* signals, with some cells exhibiting up to five *TERT* copies per cell (Figure 22). These findings showed that both telomere stabilization mechanisms, ALT and *TERT* activation, were present within the same population in LFS172.

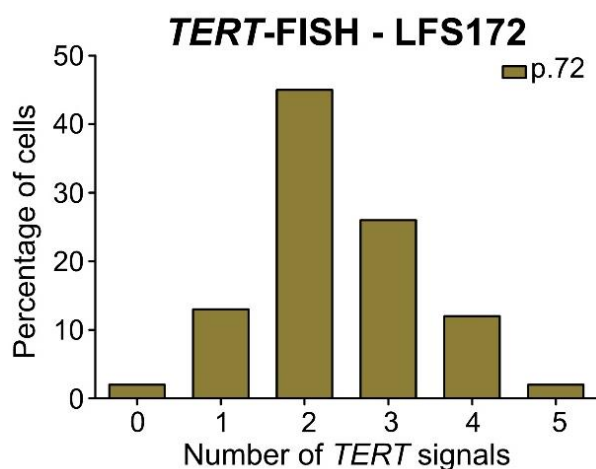


Figure 22. TERT-FISH analysis shows gains and losses in *TERT* locus in LFS172 post-crisis passage (p.72). 100 nuclei were quantified. *TERT*-FISH and quantifications were performed by Frauke Devens.

In summary, FISH results on ALT and *TERT* revealed that even in early and crisis passages, rare cells already exhibited activation of telomere maintenance mechanisms. Furthermore, the analysis demonstrated that while the majority of cells in post-crisis and late passages showed increased *TERT* (LFS041 and LFS174) or ALT (LFS087 and LFS172) signals, a significant proportion of cells did not exhibit activation of these mechanisms. These observations highlight the possibility that both ALT and *TERT* mechanisms may coexist within a single population of cells, with different subclones relying on distinct mechanisms for telomere stabilization.

3.7. Inducing cellular immortalization via *hTERT* transduction

As telomere stabilization was identified as a key event in dominant clones, I aimed to investigate whether immortalizing LFS cells via *hTERT* transduction could prevent chromothripsis or further genomic instability. For this, I collaborated with Florian Selt to perform a retroviral transduction of early-passage cells (p.18) for all four patients with human *TERT* (*hTERT*⁺). In addition, cells from the same passages were transduced with an empty pBabe-hygro vector and a pMX-GFP vector, giving us two negative controls. Unfortunately, all control cells (empty and pMX-GFP vectors) from patient LFS087 died shortly after transduction. Similarly, empty vector-transduced cells in the other three patients (LFS041, LFS172, and LFS174) also exhibited complete cell death at early stages following the procedure.

Results

3.7.1. Accelerated growth and decreased stress morphology in *hTERT*⁺ cells, with no signs of crisis

Similar to what was described earlier (see Section 3.2), negative control cells entered a growth crisis, displaying increased morphological heterogeneity. These changes included altered cell morphology, enlarged and apoptotic/necrotic cells, and substantial cell debris (Figure 23A, C). In contrast, *hTERT*⁺ cells retained their typical fibroblast morphology and exhibited rapid growth directly after transduction, without showing any indication of cellular stress.

As expected, negative control cells entered crisis between passages 26–28 in LFS041 and 30–33 in LFS174. Remarkably, *hTERT*⁺ cells did not seem to enter a crisis phase, as they continued proliferating rapidly. By the end of the experiment, *hTERT*⁺ cells had reached p.45 in LFS174 (Figure 23B), p.41 in LFS041 (Figure 23D), and p.39 in LFS172 (Supplementary Figure 8) at which I stopped culturing the cells and froze the samples.

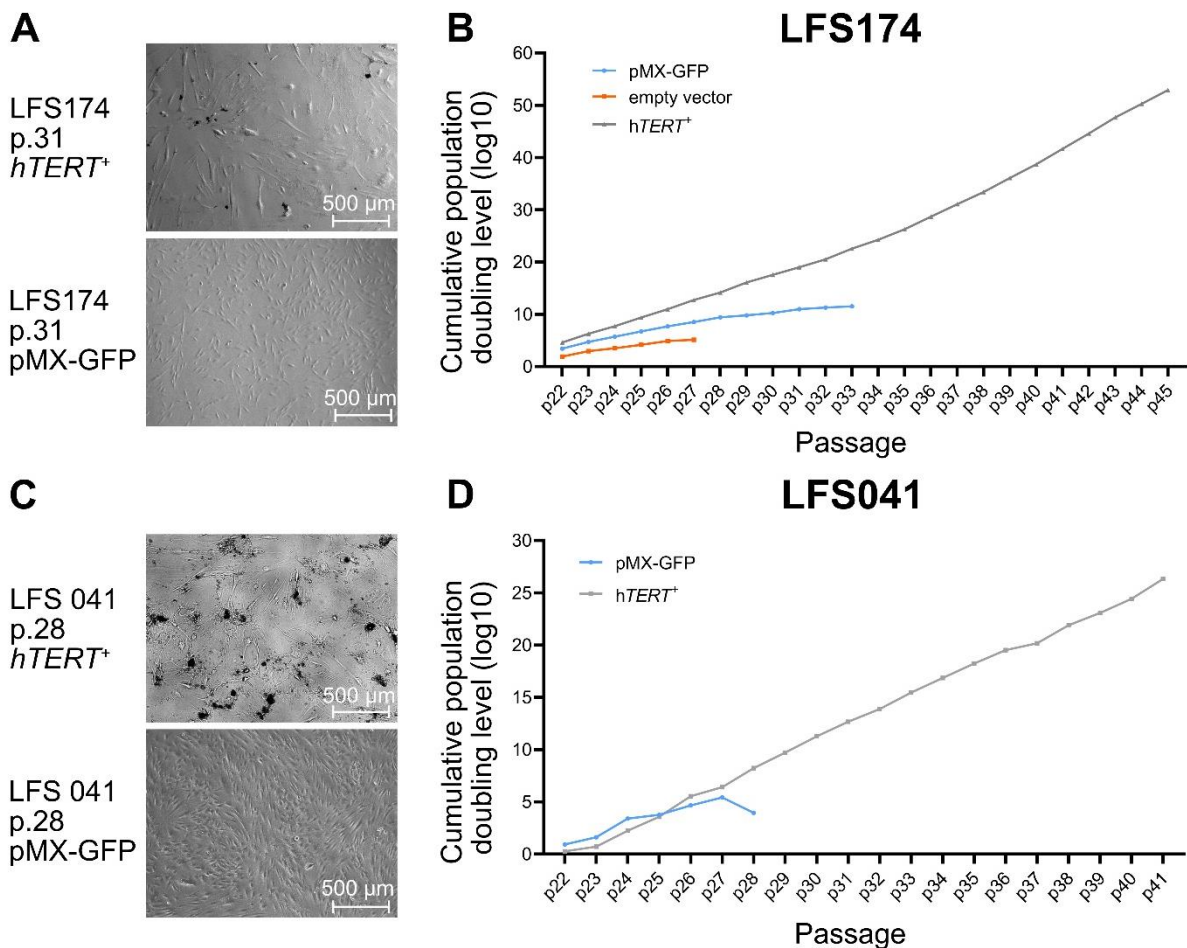


Figure 23. Morphological and growth characteristics of LFS *hTERT*⁺ and negative control cells. **A, C** Representative images depicting cellular morphology between *hTERT*⁺ and pMX-GFP control cells in LFS174 p.31 (**A**) and LFS041 p.28 (**C**). Brightfield images were acquired using a 5x magnification objective on an inverted microscope (scale bar: 500 μ m). **B, D** log10 of cumulative population doubling levels (PDLs) in LFS174 and LFS041, respectively. The graphs show the growth of *hTERT*⁺ cells (grey), along with pMX-GFP (blue) and empty vector (orange) cells. Empty vector cells in LFS041 did not survive post-transduction. GraphPad prism 8 was used to create the growth curve graphs. Retroviral transduction was performed by Florian Selt.

To quantitatively assess these observations, I calculated the cumulative population doubling level (PDL) for each passage. In LFS174, pMX-GFP cells underwent 11.5 PDLs by the time of crisis at passage 33, while *hTERT*⁺ cells achieved 24 PDLs by the same passage and 54.4 PDLs by passage 45. Similarly, in LFS041, pMX-GFP cells reached less than 4 PDLs by the crisis at passage 28, compared to 11 PDLs for *hTERT*⁺ cells at the same point and 29.1 PDLs by passage 41 (Figure 23D). These findings highlight the notable differences in morphological stability and growth dynamics introduced by *hTERT* transduction compared to negative controls across LFS cultures. Due to material availability, follow-up experiments were performed only on LFS174.

3.7.2. Significant reduction in chromatin bridges and micronuclei in LFS cells following *hTERT* transduction

For the next part, to evaluate the impact of telomerase activation on genomic stability, I quantified the frequency of chromatin bridges and micronuclei in LFS174 cells transduced with *hTERT*⁺ and pMX-GFP (control) vectors at the same passages. Matched passages from both conditions were quantified, with pMX-GFP cells being in the crisis phase. The effect of telomerase activation was pronounced in *hTERT*⁺ cells, with less than 1% of all cells exhibiting chromatin bridges, compared to 21.8% (SEM: ± 2.2) in the negative control cells (Figure 24A). Similarly, only 2% (SEM: ± 0.2) of all *hTERT*⁺ cells showed micronuclei, while micronuclei were detected in one-quarter (mean $\pm$ SEM: $25.2\% \pm 1.4\%$) of pMX-GFP control cells (Figure 24B). Together, these results demonstrated a significant reduction in chromatin bridges, rare micronuclei formation and higher genomic stability in *hTERT*⁺ cells.

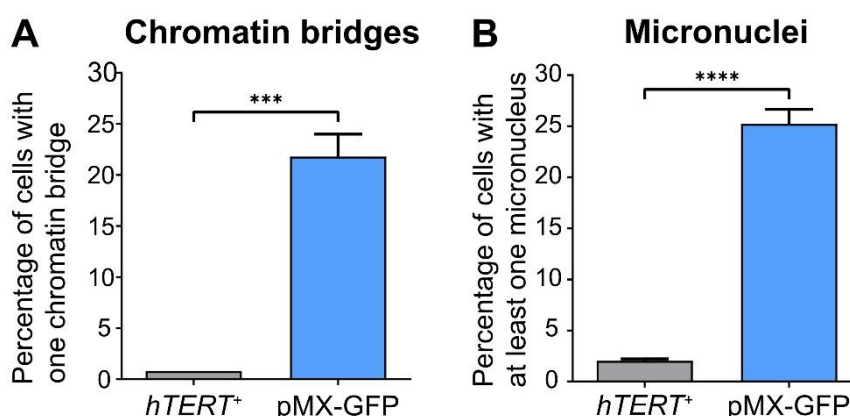


Figure 24. Quantifications of Chromatin bridges and micronuclei in LFS174 *hTERT*⁺ cells show a significant decrease compared to control cells. Bar plots show quantifications of chromatin bridges (A) and micronuclei (B) in LF174. Both *hTERT*⁺ and pMX-GFP were quantified at the same passages (p.31–p.32). Three independent biological replicates were performed for each condition, with 1000 cells quantified per replicate. Data are presented as mean $\pm$ SEM. Statistical significance was assessed using unpaired t-test. p-values lower than 0.05 were considered statistically significant. (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

Results

3.7.3. Single-cell sequencing analysis reveals reduced genomic instability in *hTERT*⁺ immortalized cells

In order to investigate genomic variations at the single-cell level and determine whether *hTERT* transduction at early passages led to maintain genome integrity and potentially prevented chromothripsis, I performed Primary Template-directed Amplification (PTA), a plate-based scDNA sequencing technique. This method offers uniform genome coverage while minimizing amplification artifacts and allelic dropouts¹³⁶. I performed the sequencing on LFS174 cells at passage 31 for both *hTERT*⁺ and pMX-GFP control cells, which represents the crisis time point for the control cells. Bioinformatics analysis was performed by Jan Otoničar.

On average, LFS control cells exhibited a markedly higher level of genomic instability, with a mean of 38.2% of the genome altered per cell for the control cells compared to 24.5% for *hTERT*⁺ cells (Figure 25).

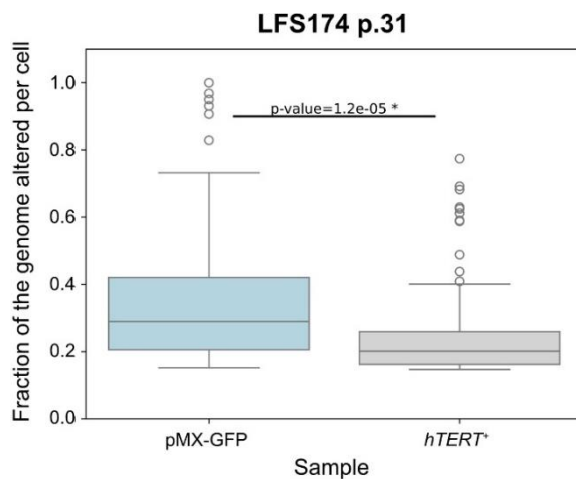


Figure 25. Fraction of the genome altered per cell in LFS174 transduced cells using PTA. p.31 was sequenced for both conditions *hTERT*⁺ (n = 112 cells) and pMX-GFP (n = 44 cells) transduced cells. Data are presented as median with interquartile range (IQR). Whiskers extend to the minimum and maximum values within 1.5x IQR, and outliers are represented as individual points. Statistical significance was assessed using the Wilcoxon rank-sum test. p-values less than 0.05 were considered statistically significant (*p < 0.05). Jan Otoničar performed the bioinformatics analysis.

Copy number plot analysis revealed shared CNAs between *hTERT*⁺ and control cells. These shared events included clonal losses of 3p, 4q, 6p, 9p, 13, 17p and CN gain of 18p (Figure 26). Given that LFS174 is derived from a female patient, the Y chromosome is not shown on the heatmap¹²².

Interestingly, control cells exhibited additional clonal CNAs, which were absent in *hTERT*⁺ cells. The most striking feature in control cells was the clonal loss of 8p and Xp, which was not observed in *hTERT*⁺ cells. Furthermore, control cells frequently displayed high-level gains (≥8 copies) across multiple chromosomes, whereas such events were rare in *hTERT*⁺ cells.

A subset of control cells appeared to be near haploid, which I also did not observe in *hTERT*⁺ cells. However, this observation requires further validation through metaphase spread analysis to exclude potential technical artifacts, such as low sequencing coverage or high allele dropout rates. In contrast, *hTERT*⁺ cells did not exhibit distinct clonal CNAs.

Copy number analysis also revealed subclonal events in both samples, highlighting the genomic heterogeneity within the populations. These CNAs included 1p loss, complete chromosome 22 loss, and telomeric 6p loss in control cells, while *hTERT*⁺ cells exhibited partial chromosome 22 loss and a 3p gain.

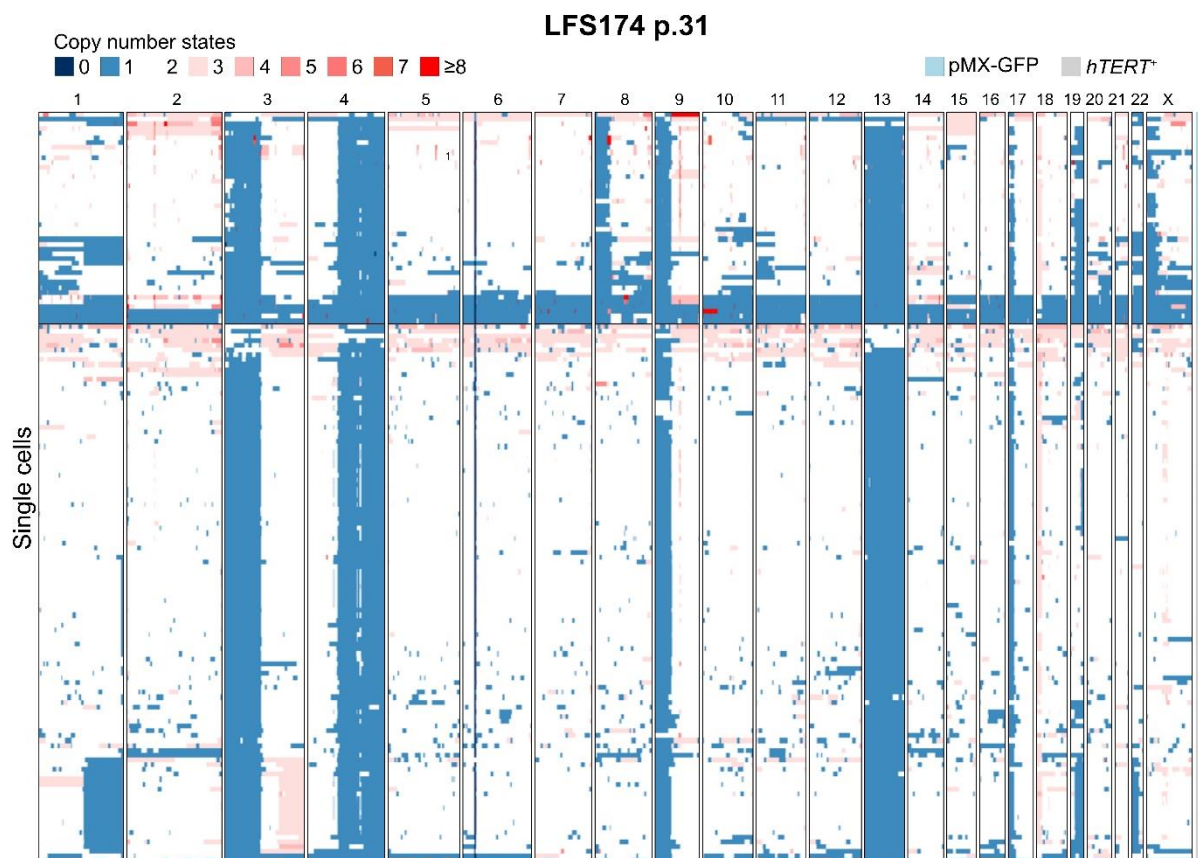


Figure 26. Copy number analysis for LFS174 *hTERT* and pMX-GFP control cells at passage 31 from PTA sequencing. Matched passages for both conditions were sequenced and analysed using scAbsolute. Each row represents one cell, and each column represents one chromosome with a bin size of 500 kb. Jan Otoničar performed the bioinformatics analysis.

Together, these findings revealed that pMX-GFP control cells exhibited markedly higher genomic instability than *hTERT*⁺ cells, characterized by increased CNAs, aneuploidy, and a higher prevalence of cells with high copy number gains. Telomerase activation in early passages via *hTERT* transduction appeared to mitigate genomic alterations, likely playing a protective role in maintaining chromosomal integrity in LFS cells.

3.8. Drivers of dominant clones

After examining the role of telomere stabilization in conferring chromothriptic cells a selective advantage, as well as identifying key driver genes using bulk WGS, I sought to explore additional essential features leading to the clonal dominance. Specifically, I focused on ecDNAs and gene fusions, as they are strongly linked to chromothripsis^{20,21,32,88}. I aimed to elucidate their timing and the specific driver genes implicated in clonal selection, which are crucial for understanding the mechanisms underlying clonal dominance.

Results

3.8.1. ecDNAs as a defining feature of winning clones

Quantification of ecDNA in metaphase spreads revealed dynamic changes across passages. In LFS041, I detected ecDNAs in one quarter of early-passage cells (p.19), with a maximum of 6 copies per cell (Figure 27B). By the crisis phase (p.27), nearly half of the cells contained ecDNAs, ranging from 0 to 27 copies per cell. The proportion of ecDNA-positive decreased to almost one third in the post-crisis passage (p.65) and further declined in the late passage (p.348), where only 9% of cells retained ecDNAs.

In LFS087, early-passage cells (p.23) exhibited minimal ecDNA-positive cells (2%). However, similar to LFS041, I observed a markable increase during the crisis phase (p.33), with nearly 38% of cells harbouring ecDNAs. At post-crisis passages (p.37, p.47), the prevalence of ecDNA-positive cells declined. Notably, p.81 showed the highest proportion of cells with more than 5 ecDNAs per cell (22%), before further decreasing in the late passage (p.192).

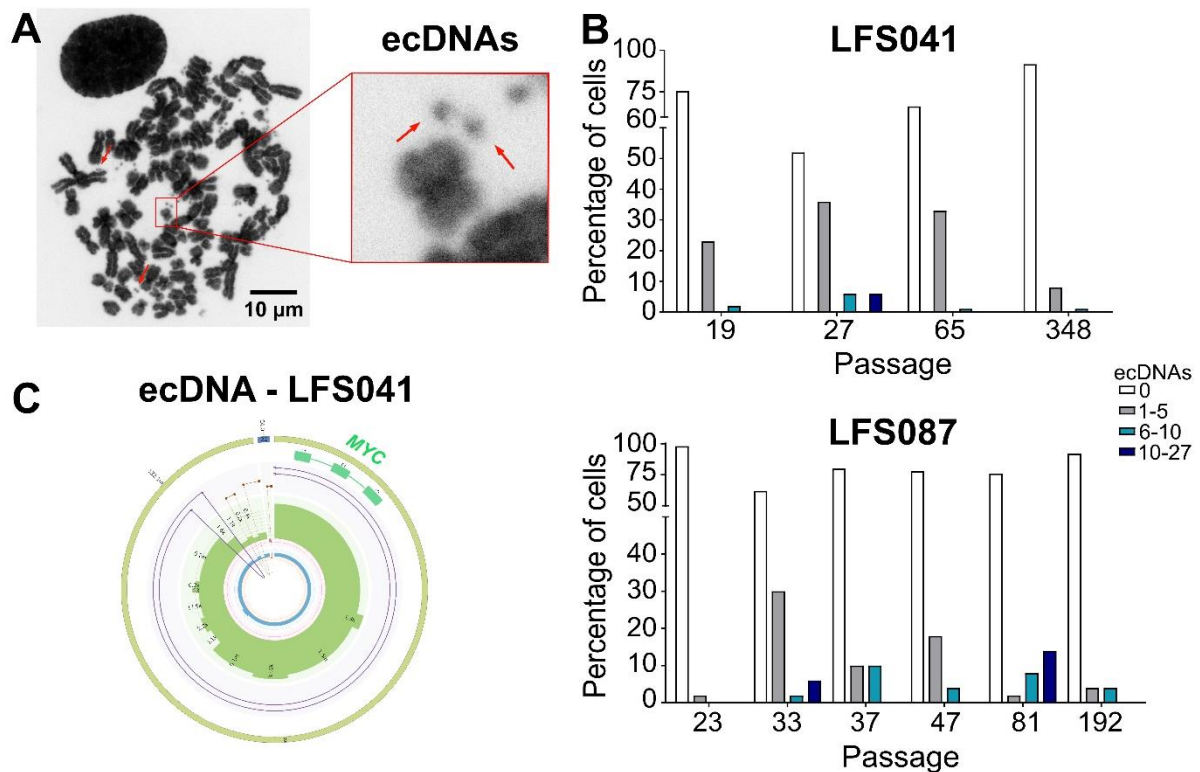


Figure 27. Quantification and analysis of ecDNA in LFS041 and LFS087 Cells. **A.** Representative image of metaphase spreads showing high numbers of ecDNAs. Scale bar: 10 µm. **B.** Quantification of the percentage of cells harbouring ecDNAs in LFS041 and LFS087 across multiple passages. For LFS041, time points include early (p.19), crisis (p.27), post-crisis (p.65) and late (p.348) passages. For LFS087, time points include early (p.23), crisis (p.33), post-crisis (p.37, p.47) and late (p.81, p.192) passages. A minimum of 50 cells per passage were quantified for each patient. **C.** Example of ecDNA structure detected in LFS041 p.63 using bulk WGS containing the oncogene *MYC*. ecDNAs were identified using AmpliconSuite, which was performed by Michelle Chan. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Frauke Devens performed DAPI staining and ecDNA quantification, while Petr Smirnov ran OncoAnlyzer.

I next looked at ecDNA structures from bulk WGS on LFS041 p.63 and LFS087 p.195. For this, Petr Smirnov ran OncoAnalyzer, and Michelle Chan conducted the AmpliconArchitect analysis.

The analysis revealed the presence of ecDNAs in both samples. In LFS041, three ecDNAs originating from chromosomes 6, 8, and 10 (Supplementary Figure 9) were identified. Notably, the ecDNA derived from chromosome 8 contained the *MYC* oncogene and the *PVT1* gene. The presence of *MYC* was further supported by OncoAnalyser data (Figure 27C). In LFS087, a single ecDNA was detected on chromosome 12, but no genes were detected on that amplicon. Collectively, these results demonstrated that both cultures exhibited a marked increase in ecDNA prevalence from early to crisis phases, followed by a decline in later passages, with *MYC* emerging as a key driver of clonal dominance in LFS041.

3.8.2. Gene fusion detection through bulk genomic and transcriptomic analysis

Beyond ecDNAs, recent research has shown that chromothripsis plays a critical role in driving oncogenesis by generating gene fusions that disrupt tumour suppressor genes or activate oncogenes^{24,115}. Based on bulk RNA sequencing (RNA-seq), for which Frauke Devens performed the RNA extractions, I detected a gradual increase in the number of gene fusions in both LFS041 and LFS087. By quantifying high-confidence gene fusions, I found that their number was minimal in early and crisis passages for both patients. In contrast, post-crisis and late passages showed a marked increase in gene fusions (Figure 28).

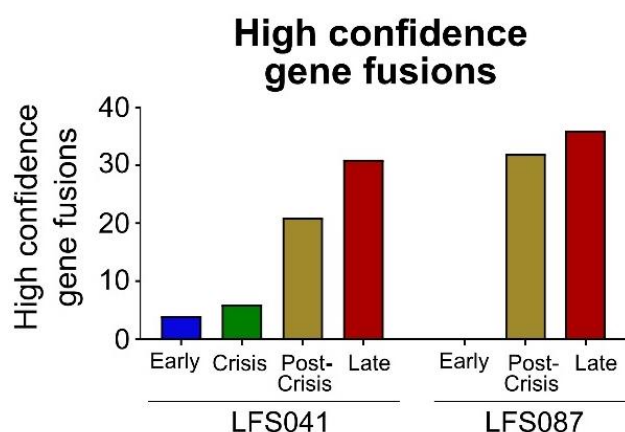


Figure 28. High-confidence gene fusions identified from bulk RNA-seq. Four passages from LFS041 were sequenced: early (p.19), crisis (p.29), post-crisis (p.65), and late (p.346), as well as three passages from LFS087: early (p.20), post-crisis (p.47), and late (p.192). Only high-confidence fusions are shown. Gene fusions were identified through Arriba analysis of bulk RNA-seq data. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Frauke Devens performed RNA extractions.

Through bulk RNA-seq analysis, I identified several candidate fusions exhibiting notable differential expression between passages. Among these, multiple *TRIM40* fusions were detected in the late passage of LFS087 (p.192), including *TRIM40-RFX8*, *TRIM40-ADM*, *TRIM40-COL6A3* and *TRIM40-C7orf50*. Notably, *TRIM40* showed a substantial increase in expression, with an FPKM value of 124.7 in the late passage, but it was barely detectable (FPKM < 0.01) in earlier passages (Figure 29 and Supplementary Figure 10). Similarly, *ERC1* fusions involving *WNT5B* and *CACNA1C* were detected in LFS087 p.192, whereas *ERC1* expression increased by nearly 2.5-fold in late passage compared to early passage (Supplementary Figure 11). In LFS041, I identified *COL5A1-SCAMP2* fusion in LFS041 late passage (p.346), with *COL5A1* expression showing a 10-fold increase compared to the early passage (p.19) (Supplementary Figure 12).

Results

Moreover, analysis of bulk WGS data using OncoAnalyser revealed 67 gene fusions in LFS087 late passage and 23 gene fusions in LFS041 post-crisis. Interestingly, five gene fusions were shared between bulk RNA-seq and WGS data in LFS087 late passage, which included *STX5-WDR74*, *NUP214-PRRC2B*, *BMPR1A-MTMR10*, *TANGO6-CDH3* and *CACNA1C-ERC1*. In LFS041 post-crisis, two gene fusions, *PC-STAR10* and *KDM2A-TCIRG1*, were shared between RNA-seq and WGS data. Thus, a significant increase in gene fusions in post-crisis and late passages of LFS041 and LFS087 was revealed through bulk RNA-seq and WGS analysis, with several candidate fusions exhibiting notable differential expression.

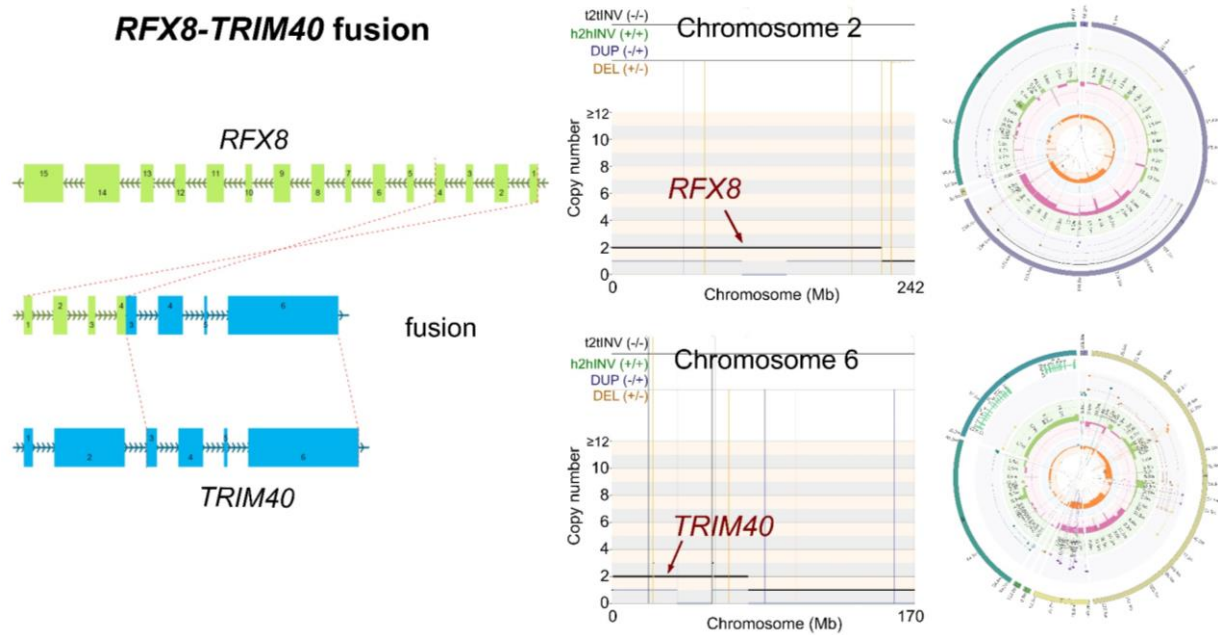


Figure 29. Representative example of the RFX8-TRIM40 gene fusion in LFS087 p.195 (late passage), identified with bulk RNA-seq. Left panel: Breakpoint sites of the *RFX8* (chromosome 2) and *TRIM40* (chromosome 6) genes involved in the fusion, which was identified through Arriba analysis of bulk RNA-seq data. Middle panels: ReConPlot of chromosomes 2 and 6 from bulk WGS of LFS087 p.195 (late passage), highlighting *RFX8* and *TRIM40*, respectively, and showing SVs on both chromosomes. Colour coding indicates deletions (orange), duplications (blue), head-to-head inversions (green), and tail-to-tail inversions (black). Total chromosome length is shown in megabase (Mb). Abbreviations: t2tINV: tail-to-tail inversion, h2hINV: head-to-head inversion, DUP: duplication, DEL: deletion. Right panel: Circos plots from bulk WGS of LFS087 p.195 showing SVs supporting the gene fusion detected by RNA-seq. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Petr Smirnov performed the bioinformatics analysis of bulk WGS.

3.9. Temporal profiling using mass spectrometry-based proteomics

To investigate the mechanisms and signalling pathways that drive chromothripsis and their contribution to the clonal selection, I collaborated with Gianluca Sigismondo to perform mass spectrometry-based proteomics. The experiment was conducted at three distinct time points in patient LFS041: early (p.19), crisis (p.27), and late (p.346), with each condition analysed using two biological replicates to ensure reliability and reproducibility (see Methods 6.2.12).

Hierarchical clustering of proteins, based on their Z-scores, revealed a pronounced upregulation of proteins in the late-passage cells, in stark contrast to the early-passage cells, which predominantly exhibited downregulated proteins (Figure 30A). The crisis passage, however, displayed a more heterogeneous profile, with both upregulated and downregulated proteins, as well as proteins with minimal dysregulation. Furthermore, the analysis identified distinct protein expression patterns across the three passages. Some protein clusters were upregulated in two passages but downregulated in the third, while other clusters showed the opposite pattern, being downregulated in two passages and upregulated in the remaining one, highlighting passage-specific dynamics.

In pairwise comparisons of protein levels, the crisis phase demonstrated a nearly 20-fold increase in significantly upregulated proteins (3599 proteins) compared to the early passage (175 proteins) (Figure 30B). Similarly, a comparison of early and late passages demonstrated a marked increase in upregulated proteins in the late passage, with 4456 proteins showing significantly higher expression compared to 578 upregulated proteins in the early passage (Figure 30C). While both crisis and late passages showed a substantial increase in protein levels, the late passage exhibited more extensive upregulation, with 1887 proteins upregulated – more than double the 845 proteins in the crisis (Figure 30D).

Together, these results provided evidence of significant proteomic remodelling throughout cellular progression, with clear distinctions in protein levels between the early, crisis, and late passages.

3.9.1. Gene set enrichment and pathway analysis of differentially expressed proteins

To gain insights into the biological processes and pathways associated with the observed proteomic changes, I performed gene set enrichment analysis (GSEA). This analysis encompassed three complementary approaches: gene ontology (GO), KEGG (Kyoto Encyclopedia of Genes and Genomes), and Reactome. GO analysis offers a broad overview of biological processes, molecular functions, and cellular components, while KEGG and Reactome provide pathway-specific insights. Furthermore, to visualize and summarize the relationships between enriched pathways, I used Cytoscape's EnrichmentMap¹³⁷ and AutoAnnotate¹³⁸ clustering tools to identify and annotate key biological processes underlying the proteomic changes.

3.9.1.1. Identification of key biological processes using GSEA GO

The GSEA GO analysis comparing early and crisis stages revealed a predominant upregulated enrichment in a number of processes during the crisis, particularly those related to metabolic biosynthesis and nucleotide-related processes (Figure 31A).

Results

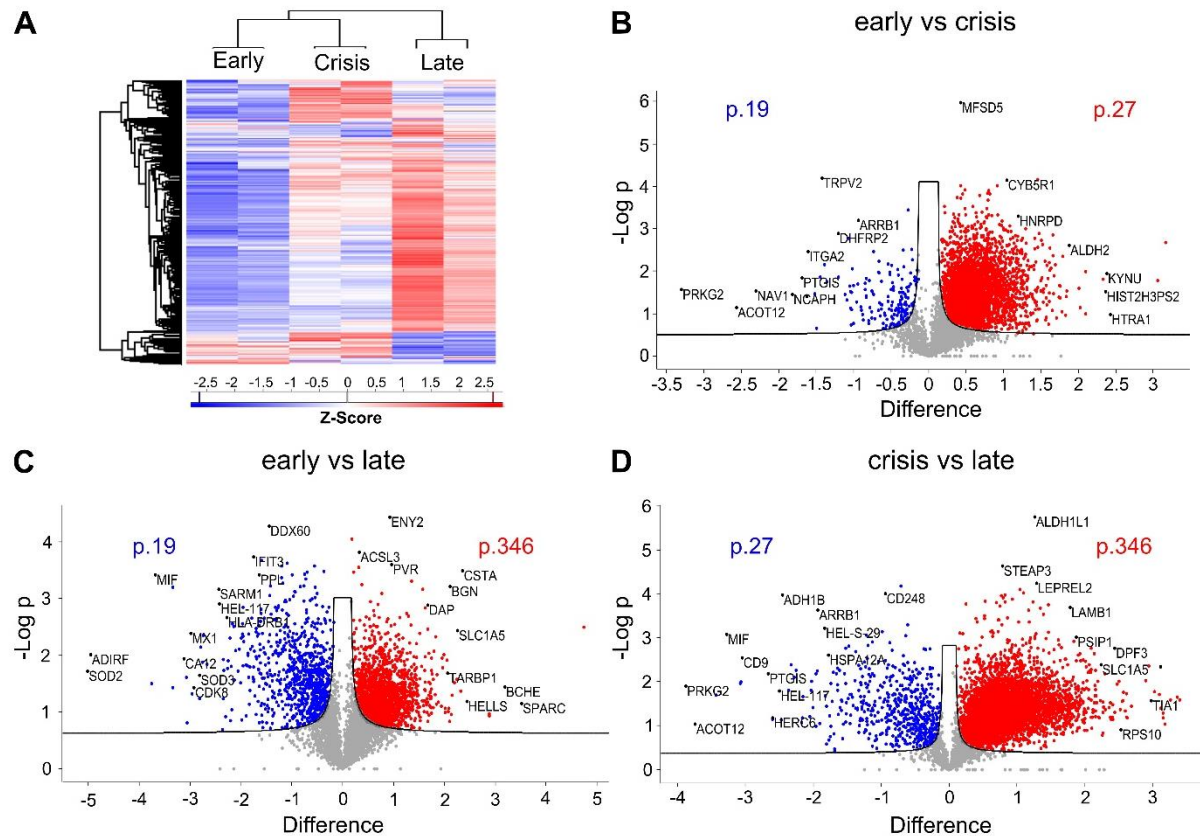


Figure 30. Differentially regulated proteins across cellular passages revealed by mass spectrometry-based proteomics. **A.** Hierarchical clustering of proteins based on Z-scores, showing distinct patterns in early (p.19), crisis (p.27), and late (p.346) passages from patient LFS041. Proteins are grouped into clusters based on their upregulation (red) or downregulation (blue) across the three passages. **B.** Volcano plot comparing protein expression between early (p.19) and crisis (p.27) passages, highlighting significantly upregulated proteins (red) and downregulated proteins (blue). The x-axis represents the log2 fold change (log2FC) in protein expression, calculated as the difference in average log2-transformed intensity values between the two passages, using two biological replicates per passage. The y-axis represents the statistical significance ($-\log_{10}$ p-value). **C, D** Volcano plots comparing early (p.19) versus (vs.) late (p.346) passages and crisis (p.27) vs. late (p.346) passages, respectively. Panel **A** taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Gianluca Sigismondo performed the experimental part and ran the analysis on Perseus. Plots were generated by Gianluca and myself.

Key upregulated processes included (purine-) nucleoside triphosphate biosynthesis, (purine-) nucleotide metabolism, purine nucleotide metabolism and purine ribonucleoside triphosphate metabolism, indicating an increased demand for nucleotide synthesis and metabolism during the crisis phase. Other metabolic processes, such as ATP metabolic processes, catabolic fatty acid and lipid processes, as well as oxidative phosphorylation were upregulated in the crisis, reflecting a stress response and a shift in energy production (Figure 31A, Supplementary Figure 13A). Conversely, peroxisome proliferator-activated receptor (PPAR) pathway downregulation during the crisis, along with upregulation of peptide antigen processing and MHC complex presentation, suggests a role in immune response and cellular metabolism.

A comparison of early and late passages supported my previous observations, showing an upregulation of telomere maintenance via telomerase and telomere lengthening in the late passage (Figure 31B, Supplementary Figure 13B). In contrast, the early passage showed significant downregulation of mitotic and cell cycle processes, such as spindle organization, chromosome segregation, and mitotic phase transitions, suggesting impaired checkpoint activation at the mitosis, thus leading to mitotic defects. New insights were also uncovered: helicase activity, replication fork, DNA replication and strand elongation were all downregulated in the early passage, implicating replication stress. Furthermore, I observed increased transcription and translation activity in the late passage, covering processes like transcription by RNA polymerase II, rRNA metabolic process, ribosome biosynthesis. In addition, late passages showed enhanced post-transcriptional gene regulation and DNA conformational changes, consistent with chromosomal instability associated with chromothripsis.

The enrichment analysis of crisis versus late passage provided additional perspective on the temporal changes of cellular processes (Figure 31C, Supplementary Figure 13C). Late passages exhibited upregulation of pathways related to ncRNA and mRNA metabolism, DNA metabolism, and regulation of nucleobase-containing compound metabolism, reflecting increased activity in gene expression regulation and DNA repair processes. In addition, regulation of centrosome duplication and microtubule cytoskeleton organization involved in mitosis were enriched in the late passage, indicating increased genomic instability and mitotic errors during the crisis.

Conversely, crisis passages showed downregulation of unsaturated fatty acid metabolism, carbohydrate metabolism, reactive oxygen species (ROS) metabolism, and lipid catabolism, indicative of metabolic stress. Furthermore, vesicle-mediated transport and immune response, including upregulation of MHC protein complex, immune response-activating signalling pathway, and inflammatory response were all enhanced in crisis-passage cells. Additionally, Cell at this passage showed an elevated requirement for nucleotide production, as evidenced by the upregulation of processes such as nucleoside bisphosphate catabolic process and pyridine nucleotide biosynthetic process. Notably, crisis cells demonstrated upregulation in fibroblast proliferation, cell adhesion and positive regulation of apoptotic processes.

As in the previous comparison, late passages demonstrated upregulation in processes related to telomere maintenance, mitosis, RNA splicing, DNA replication, transcription, and translation compared to crisis passages. These included telomere organization, centrosome duplication, attachment of spindle microtubules to kinetochore, chromosome segregation, helicase activity, DNA replication initiation, replication fork, DNA-templated transcription initiation and elongation, ribosome biogenesis, rRNA processing, nucleolar activity and translational initiation.

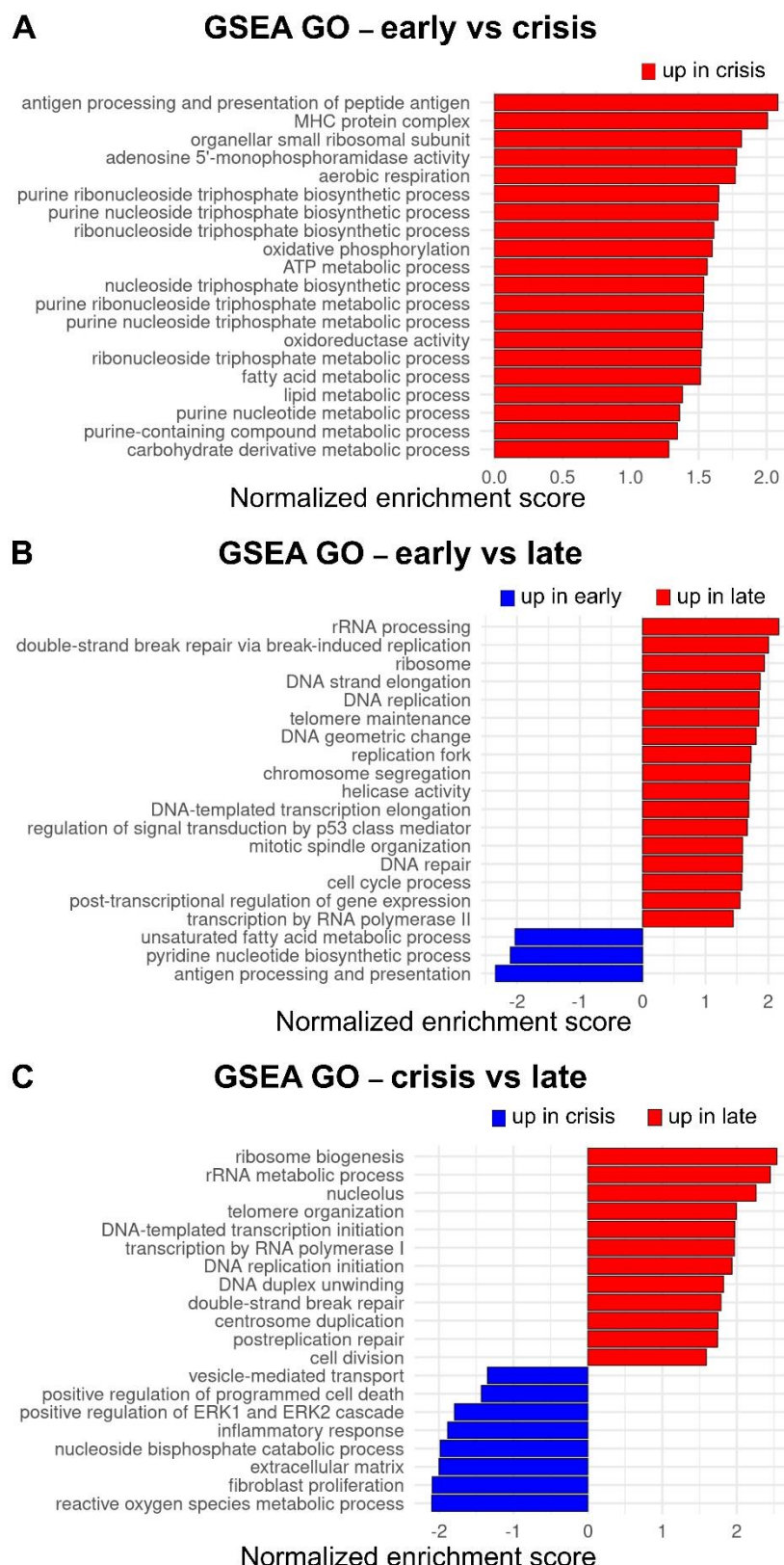


Figure 31. GSEA enrichment analysis of gene ontology (GO) in LFS041 fibroblasts using mass spectrometry-based proteomics. Selected significantly enriched GSEA GO processes across three passages in patient LFS041. Comparisons include early (p.19) vs. crisis (p.27) passages (**A**), early (p.19) vs. late (p.346) passages (**B**), and crisis (p.27) vs. late (p.346) passages (**C**). The x-axis shows the normalized enrichment score, and the y-axis displays the description of the GO processes. Panel **B** adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I).

Taken together, GSEA-GO results highlight a variety of dysregulated processes, including DNA damage repair, mitotic stress, and telomere maintenance. It also provides novel insights into mechanisms underlying chromothripsis, such as increased nucleotide demand, metabolic stress, DNA replication stress, and enhanced transcriptional and translational activities.

3.9.1.2. Complementary pathway insights with GSEA Reactome and GSEA KEGG

To gain deeper mechanistic insights at the pathway level and to explore molecular interactions, I conducted GSEA Reactome and GSEA KEGG analyses on the proteomics data.

Similar to the GSEA GO analysis, GSEA Reactome results revealed that most pathways were downregulated in the early passage compared to both crisis and late passages (Figure 32A, B, Supplementary Figure 14A, B). In contrast, the crisis passage exhibited significant upregulation of key pathways distinct from the early and late phases (Figure 32A, C, Supplementary Figure 14A, C). These pathways included pyruvate metabolism and the TCA cycle, reflecting increased energy demands. Metabolic pathways such as peptide hormone metabolism, fatty acid metabolism, and lipid metabolism were also enriched during the crisis, highlighting altered metabolic signalling. Consistent with *TP53* second hit, the crisis passage showed an upregulation of defective intrinsic apoptotic pathways compared to the early passage. Both early and crisis passages exhibited enrichment of metabolic processes and immune response pathways, such as interferon signalling and antigen presentation, when compared to late passages.

In line with the GSEA GO findings, relative to early and crisis passages, late-passage cells showed upregulation of pathways related to *TP53* signalling, DNA replication, transcription, major pathway of rRNA processing in the nucleolus and cytosol, DNA repair (NHEJ, BER, HDR), and telomere maintenance (Figure 32B, C, Supplementary Figure 14B, C). These included specific processes such as DNA unwinding, activation of the pre-replicative complex, DNA strand elongation, epigenetic regulation, mitotic spindle and cell cycle checkpoints. Pathways involving transcriptional regulation by small RNAs and generic transcription were also prominently enriched in late passages. Interestingly, I noticed a significant downregulation of SUMOylation pathways in early- and crisis-passage cells, including the SUMOylation of DNA replication proteins, transcription cofactors, chromatin organization proteins, and RNA-binding proteins, indicating a disruption in key cellular processes.

GSEA analysis of KEGG pathways supported the findings from GSEA GO and Reactome analyses (Figure 32D, E), whereas crisis-passage cells demonstrated upregulated enrichment in numerous metabolic pathways (e.g., carbon and fatty acid metabolism), antigen processing and presentation, oxidative phosphorylation and TCA cycle (Figure 32D, F).

Results

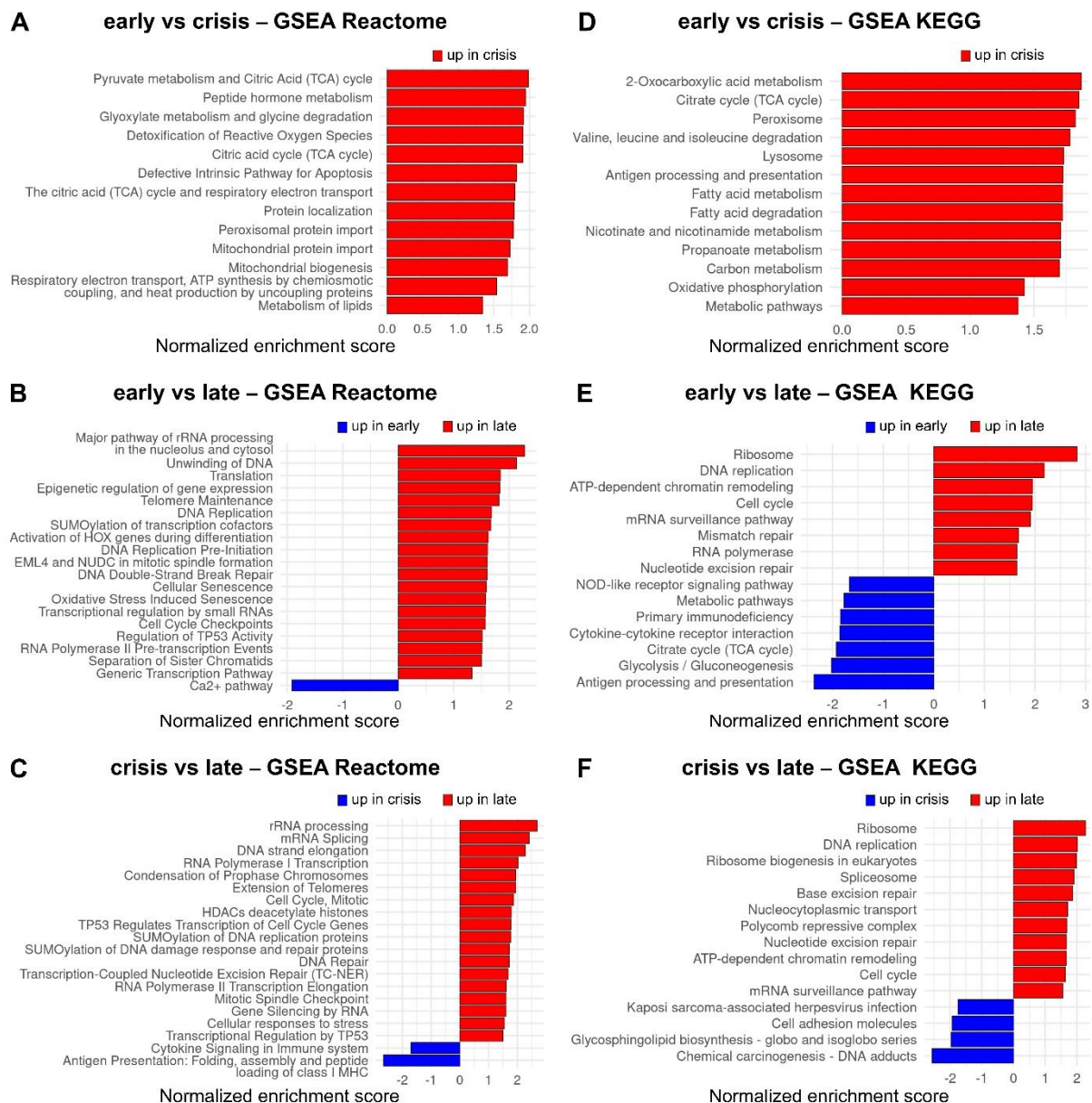


Figure 32. GSEA Reactome and KEGG pathway analysis in LFS041 proteomics data. A. B. C. Selected significantly enriched GSEA Reactome processes across three passages in patient LFS041. Comparisons include early (p.19) vs. crisis (p.27) passages (**A**), early (p.19) vs. late (p.346) passages (**B**), and crisis (p.27) vs. late (p.346) passages (**C**). The x-axis shows the normalized enrichment score, and the y-axis displays the description of the Reactome pathways. **D. E. F.** Selected significantly enriched GSEA KEGG processes across three passages in patient LFS041. Comparisons include early (p.19) vs. crisis (p.27) passages (**D**), early (p.19) vs. late (p.346) passages (**E**), and crisis (p.27) vs. late (p.346) passages (**F**). The x-axis shows the normalized enrichment score, and the y-axis displays the description of the KEGG pathways.

On the other hand, late-passage cells exhibited upregulation in pathways involved in DNA replication, ribosome-related processes, cell cycle, mRNA surveillance and DNA repair (MMR, NER, BER). Additionally, cell adhesion showed to be downregulated in late-passage cells compared to both early and crisis passages. Overall, GSEA Reactome and KEGG analyses not only support the results from GSEA GO but also provide deeper insights into the pathways underlying the observed biological processes.

3.9.1.3. Analysis of protein levels in key GSEA GO processes and pathways

Next, I investigated the differential protein expression profiles of the GSEA GO processes that I selected based on their potential roles in elucidating the molecular mechanisms underlying chromothripsis initiation.

Notably, transcription-related processes showed increased protein levels from early passage to crisis and late passages. This pattern was particularly evident for pathways such as DNA-templated transcription initiation (Figure 33), rRNA transcription, basal transcription machinery binding, and transcription mediated by RNA polymerases I and II (data not shown). The observed upregulation in transcriptional activity following the loss of the WT *TP53* allele during the crisis and in subsequent passages may explain the global upregulation of protein expression observed in these stages compared to early passage cells.

Furthermore, I analysed the protein profiles of several nucleoside and nucleotide metabolic processes, which also demonstrated significant upregulation during the crisis. Processes such as the purine nucleotide metabolic process exhibited marked upregulation in protein expression in the crisis passage, not only compared to the early passage but also to the late passage. Given this notable disruption in nucleotide metabolism, I further examined pathways and processes that influence the nucleotide pool and regulate nucleotide biosynthesis. This deeper investigation revealed dysregulation in several key proteins involved in the pentose phosphate pathway, a critical metabolic pathway for nucleotide biosynthesis¹³⁹. The dysregulation of the pentose phosphate pathway suggests a potential insufficiency in the nucleotide pool as cells progress through crisis. This insufficiency could impair the availability of nucleotides required for proper DNA replication and repair, thereby contributing to chromothripsis-associated genomic instability.

In addition, proteins involved in the ROS metabolic process exhibited distinct expression patterns, with a significant upregulation observed during the crisis passage, which was not sustained in the late passage. This ROS-related dysregulation may increase genomic instability and drive the metabolic shifts observed in nucleotide-related pathways, as well as contribute to the observed senescence and apoptotic activity.

Proteins associated with the inflammatory response (Figure 33), cytokine-mediated signalling pathway, and antigen processing and presentation (data not shown) were also upregulated during the crisis passage, further exacerbating the activated immune response. Furthermore, telomere organization proteins showed a gradual increase in expression from the crisis passage to the late passages, supporting my previous findings. Overall, these findings suggest that the impaired nucleotide metabolism, coupled with hypertranscription, replication and oxidative stress, may contribute to the increased genomic instability around the crisis passages.

Results

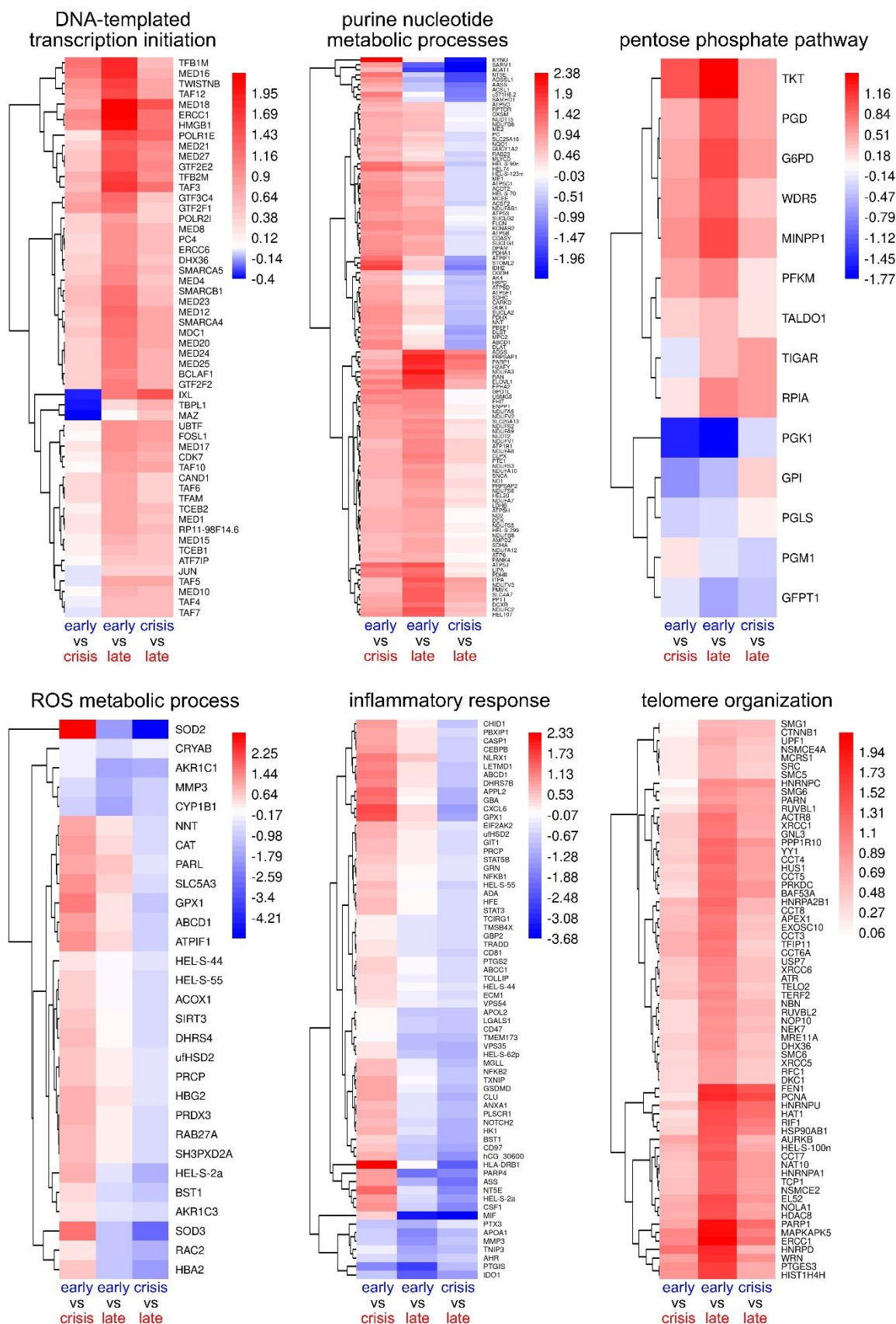


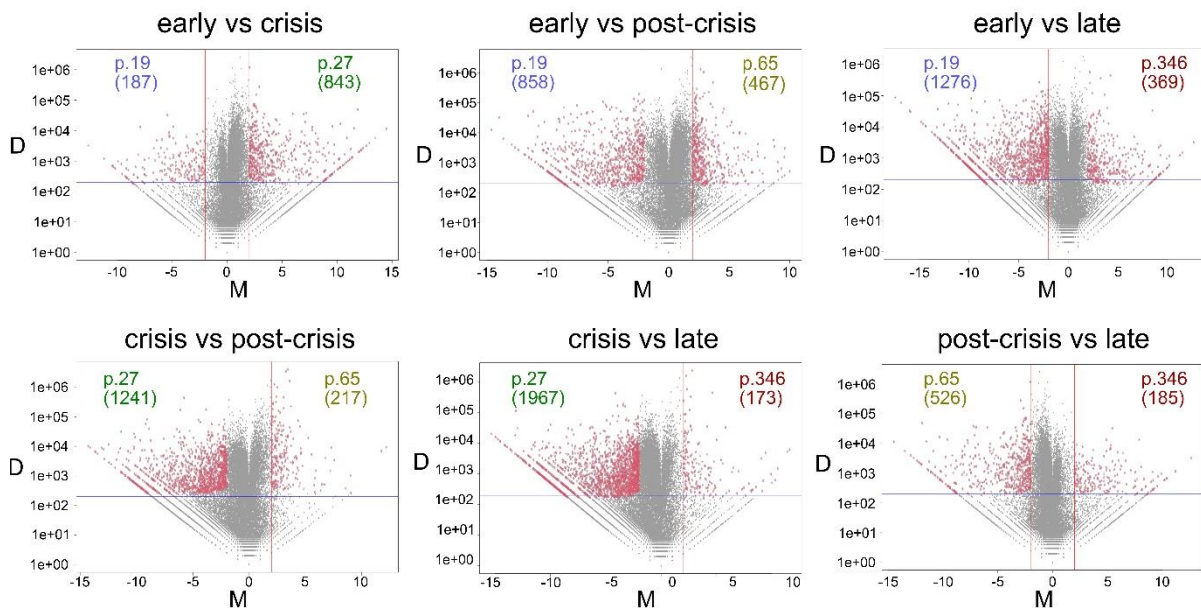
Figure 33. Differential protein expression of selected pathways and biological processes based on GSEA GO proteomics data analysis in LFS041. Each heatmap shows the differential protein expression across three comparisons: early (p.19) vs. crisis (p.27) passages, early (p.19) vs. late (p.346) passages, and crisis (p.27) vs. late (p.346) passages. The colour gradient represents the level of protein expression changes, with red indicating upregulation and blue indicating downregulation.

Results

passages. Since RNA-seq analysis includes only a single time point for the crisis and one replicate per patient, I performed the RNA-seq analysis using NOlseq^{153,154}. Unlike other differential expression methods that require multiple biological replicates, NOlseq is particularly suitable for experiments with few or no replicates, thus providing a robust approach for analysing RNA-seq data. While it is possible to use other methods (e.g. DESeq) in the absence of replicates, it results in reduced statistical power and reliability, making this approach not ideal for such analysis^{153–156}.

In LFS041, I identified a markedly higher number of significantly differentially upregulated genes in the crisis passage, not only compared to the early-passage cells, but also compared to both post-crisis and late passages (Figure 35). In contrast to the proteomics data, early-passage cells showed almost 4-fold more significantly upregulated genes relative to the late passage and nearly a 2-fold increase compared to the post-crisis passage. This pattern was also evident in LFS087, where early passage showed 603 significantly upregulated genes, as compared to 220 in post-crisis, and 727 significantly upregulated genes compared to 399 in late passages (Supplementary Figure 15).

Figure 35. Volcano plots of differentially expressed genes (DEGs) across cellular passages from



bulk RNA-seq of LFS041. The volcano plots visualize the DEGs (highlighted with a red circle) from LFS041 at four passages: early (p.19), crisis (p.29), post-crisis (p.65), and late (p.346). I performed the analysis using NOlseq. The x-axis (M) represents the log-fold change of expression levels between two conditions, indicating the magnitude and direction of change, with positive values corresponding to upregulated genes and negative values to downregulated genes. The y-axis (D) shows the absolute expression difference between conditions. Number of upregulated genes at each passage is indicated in parentheses on each plot.

3.10.1. Investigation of critical GO terms from bulk RNA-seq associated with chromothripsis and clonal expansion

Consistent with the proteomics results, GSEA GO analysis of LFS041 Bulk RNA data revealed that early-passage cells exhibited upregulation of cellular senescence and apoptotic processes compared to all other passages (Figure 36A-C and Supplementary Figure 16A-C). Processes related to immune response, such as antigen receptor-mediated signalling pathways, antigen processing and presentation, and MHC protein complex activity, were upregulated during the crisis passage compared to the early passage. However, these processes were also enriched in both early and crisis passages when compared to both post-crisis and late passages.

Furthermore, nucleoside and nucleotide metabolic processes were significantly upregulated in early and crisis passages compared to later passages. These included cyclic-nucleotide-mediated signalling, purine-containing compound catabolic processes, guanyl nucleotide binding, cytidine deaminase activity, regulation of ribonuclease activity, and ribonucleoside catabolic processes. These findings support the proteomics data, which indicated impaired nucleotide metabolism, suggesting that these impairments begin at the early passages and increase towards the crisis phase.

Markedly, post-crisis and late passages exhibited downregulation in numerous signalling pathways compared to crisis and earlier passages. These included toll-like receptor signalling pathway, vascular endothelial growth factor signalling pathway, cytokine-mediated signalling pathway, I-kappaB kinase/NF-kappaB signalling pathway, receptor signalling pathway via JAK-STAT, regulation of Ras protein signal transduction, regulation of JNK cascade, ERK1 and ERK2 cascade and positive regulation of stress-activated MAPK cascade.

In line with the GSEA GO proteomics findings, post-crisis and late passages showed increased telomerase activity, compared to early and crisis passages. Additionally, they exhibited enhanced RNA polymerase activity and upregulation of processes related to the cytosolic large ribosomal subunit and structural ribosomal components. Notably, early-passage cells displayed reduced DNA-binding transcription activator activity compared to crisis passage. However, both early and crisis passages showed upregulation of transcription-related processes, such as DNA-binding transcription factor activity, when compared to post-crisis and late passages. This suggests enhanced transcription of specific genes, which might explain the observed enrichment of upregulated signalling cascades in early and crisis passages^{157–159}. This may also indicate cellular adaptation to stress, metabolic changes, or transitions between different cellular states¹⁶⁰.

Lastly, GSEA analysis comparing post-crisis and late passages revealed fewer significantly enriched pathways, consistent with the expected similarity between these two stages. However, late passages consistently displayed downregulation of pathways, such as the Wnt

Results

signalling pathway and the regulation of transmembrane receptor protein serine/threonine kinase signalling compared to all other passages. This might correlate with the ongoing genomic alterations observed in the late-passage cells.

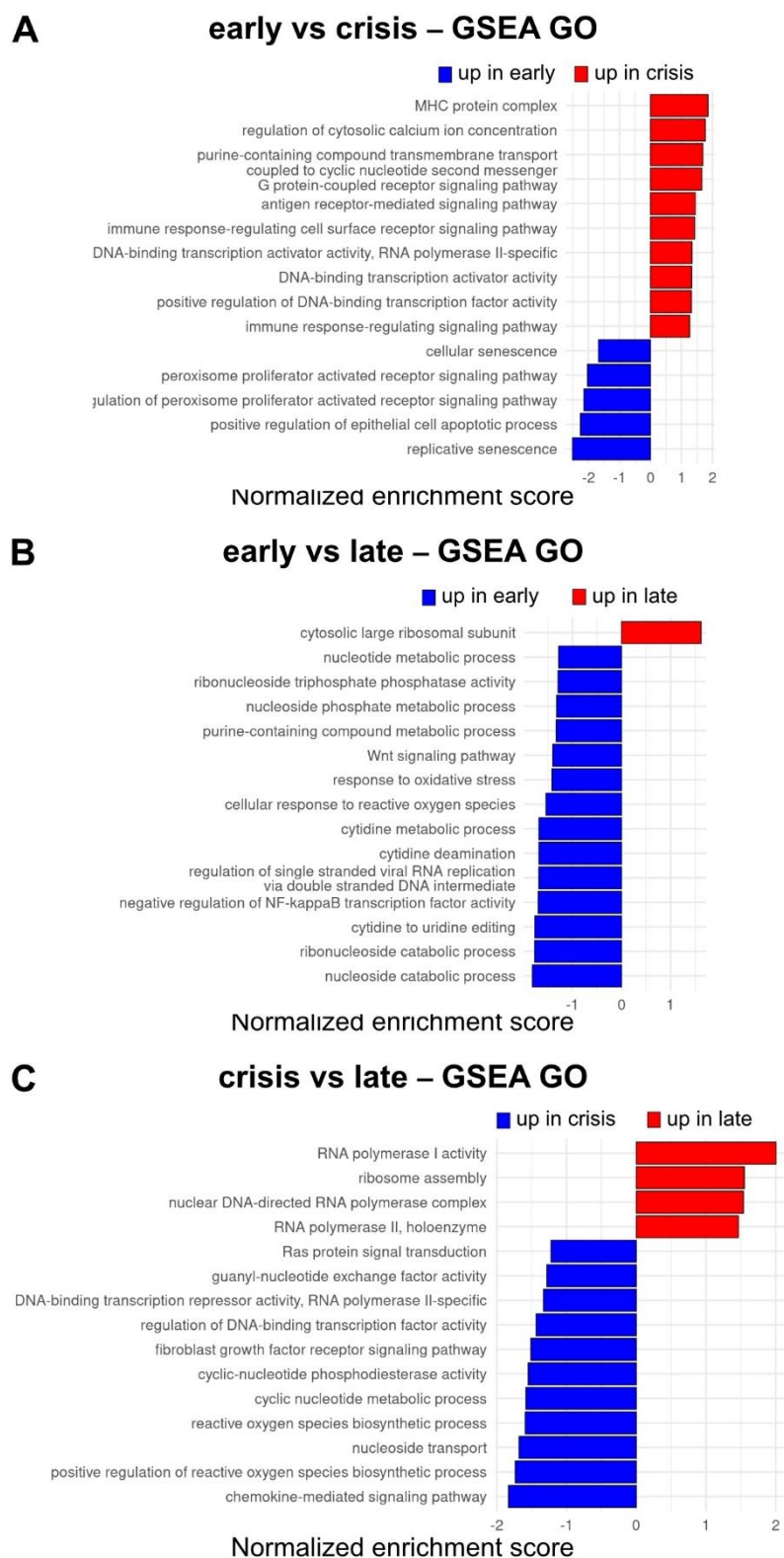


Figure 36. GSEA GO analysis in LFS041 bulk RNA-seq data. Selected significantly enriched GSEA GO processes across three passages in patient LFS041. Comparisons include early (p.19) vs. crisis (p.29) passages (A), early (p.19) vs. late (p.346) passages (B), and crisis (p.29) vs. late (p.346)

passages **(C)**. The x-axis shows the normalized enrichment score, and the y-axis displays the description of the GO terms. Comparisons including post-crisis passage are shown in Supplementary Figure 16. Panel **A** and **B** adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I).

The findings from LFS087 (data not shown) align closely with the observations reported for LFS041 and the proteomics results. Early-passage cells exhibited upregulation of several key processes, including the regulation of DNA-binding transcription factor activity, ribonuclease activity, cell adhesion, exocytosis, immune response, response to oxidative stress, Wnt signalling pathway, receptor signalling via JAK-STAT, MAPK cascade, and Ras protein signal transduction, compared to later passages.

Conversely, both post-crisis and late passages showed an enrichment in rRNA processing and ribosome assembly. However, certain processes were specifically upregulated in post-crisis cells, but not in late-passage cells, compared to the early passage. These included chromatin remodelling, DNA helicase activity, DNA strand elongation, RNA metabolic processes, and telomere organization. GSEA analysis comparing post-crisis and late passages did not identify distinctly enriched gene sets of high relevance, as expected, given the similarity between these stages.

Together, GSEA GO analysis from both patients complemented the earlier proteomics findings, highlighting consistency between RNA and protein-level data. Moreover, the bulk RNA results also provided additional insights into novel processes and pathways, offering a more comprehensive understanding of the underlying mechanisms.

3.10.2. Expanded GSEA pathway analysis using Reactome and KEGG databases

Following a similar approach to the proteomics analysis, I performed GSEA Reactome and KEGG analysis on bulk RNA-seq data from both LFS041 and LFS087 patients. For LFS041, the GSEA Reactome results exhibited similar enrichment patterns to the proteomics analysis across passages. Significantly enriched pathways included RNA polymerase III transcription initiation, SUMOylation of transcription factors, ribosome-related processes, rRNA processing, translation, mRNA splicing, cell death signalling, MAP kinase cascade, signalling by EGFR in cancer, and interferon signalling (Figure 37A-C and Supplementary Figure 17A-C). In addition to the signalling pathways revealed by GSEA GO, Reactome analysis further uncovered increased PI3K/AKT and PD-1 signalling in early and crisis passages compared to post-crisis and late passages.

In the case of LFS087, GSEA results provided additional insights that further supported previous findings from the proteomics data, some of which were not reflected in the RNA-seq data of LFS041. Notably, pathways involved in mitosis, G2/M checkpoints, mitotic prophase, condensation of prophase chromosomes, extension of telomeres, DNA replication pre-

Results

initiation, transcriptional regulation, mRNA splicing, epigenetic regulation of gene expression, DNA repair (HDR, BER), and RNA polymerase I and II transcription were all upregulated in post-crisis and late passages compared to the early passage (data not shown).

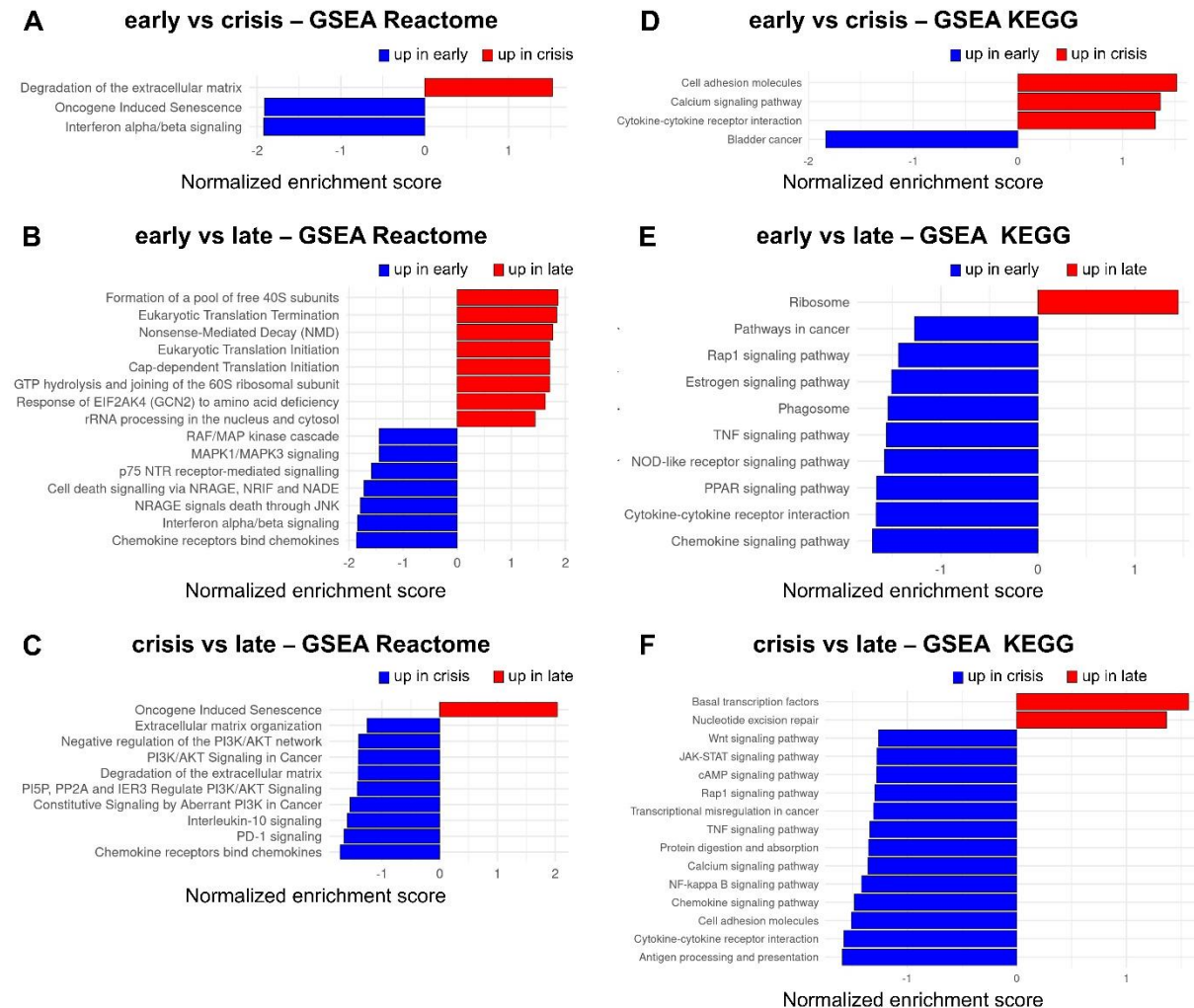


Figure 37. GSEA Reactome and KEGG pathway analysis in LFS041 bulk RNA-seq data. A. B. C. Selected significantly enriched GSEA Reactome processes across three passages in patient LFS041. Comparisons include early (p.19) vs. crisis (p.29) passages (**A**), early (p.19) vs. late (p.346) passages (**B**), and crisis (p.29) vs. late (p.346) passages (**C**). The x-axis shows the normalized enrichment score, and the y-axis displays the description of the Reactome pathways. **D. E. F.** Selected significantly enriched GSEA KEGG processes across three passages in patient LFS041. Comparisons include early (p.19) vs. crisis (p.29) passages (**D**), early (p.19) vs. late (p.346) passages (**E**), and crisis (p.29) vs. late (p.346) passages (**F**). The x-axis shows the normalized enrichment score, and the y-axis displays the description of the KEGG pathways. Comparisons including post-crisis passage are shown in Supplementary Figure 17.

Interestingly, I observed a downregulation of ATR activation in the early passage, providing deeper insight into the molecular mechanisms underlying replication stress in LFS cells. Additionally, the assembly of the origin recognition complex (ORC) at the replication origin was downregulated in the early passages, which plays an important role in DNA replication initiation and origin firing^{79,161}.

GSEA KEGG analysis further corroborated the findings from both RNA and proteomics data, while also identifying distinct pathways that were positively enriched in the early and crisis passages. Notably, pathways such as "pathways in cancer" and "transcriptional misregulation in cancer" were already significantly enriched during these passages, further confirming the activation of cancer-related pathways in the early stages (Figure 37D-F and Supplementary Figure 17D-F). Additionally, upregulation of the cGMP-PKG signalling pathway and the cytosolic DNA-sensing pathway was observed, indicating potential activation of stress response mechanisms and immune-related pathways during early and crisis passages. In conclusion, the comprehensive GSEA analysis of bulk RNA data using both Reactome and KEGG pathways not only validated the proteomics results but also offered new insights and deeper understanding into the complex interplay of signalling pathways and cellular processes.

3.10.3. Exploring DEG patterns across key GSEA GO processes and pathways

To complete the comprehensive analysis of proteomics and bulk RNA-seq data, I examined the differential gene expression patterns of the most prominent biological processes and pathways identified through GSEA of bulk RNA-seq data. Additionally, I investigated the expression of key regulatory genes for distinct pathways and processes that were identified in collaboration with Kerem's lab as potentially playing a pivotal role in the underlying mechanism. To obtain a clear overview of gene expression dynamics, I mapped the differential gene expression of LFS041 across crisis, post-crisis, and late passages, using early passage as the baseline for each comparison (Figure 38).

Remarkably, early-passage cells exhibited specific transcriptional activation compared to post-crisis and late passages, as evidenced by the upregulation of genes involved in ribosome biogenesis and DNA-binding transcription activator activity (RNA polymerase II-specific). This transcriptional activation peaked during the crisis phase, likely as a response to replication stress and genomic instability, driving the expression of survival and stress-related genes. The heightened transcriptional and translational activity during crisis was further reflected in the dysregulation of nucleotide and nucleoside metabolic processes, pointing towards an impaired nucleotide metabolism. In this context, genes involved in the pentose phosphate pathway, such as *G6PD* (glucose-6-phosphate dehydrogenase) and *PGLS* (6-phosphogluconolactonase), were downregulated, indicating a potential decreased nucleotide production and thus, deficiency in the nucleotide pool.

I also observed a notable gradual increase in replication stress from early passage to crisis, compared to later passages, with a pronounced upregulation of genes involved in DNA replication initiation, as well as DNA duplex unwinding. During early and crisis passages, genes associated with chromosome segregation (data not shown) and mitotic spindle organization were upregulated, indicating increased mitotic abnormalities. This upregulation aligns with the

Results

formation of multipolar spindles, which could drive chromosomal missegregation and aneuploidy, as previously observed.

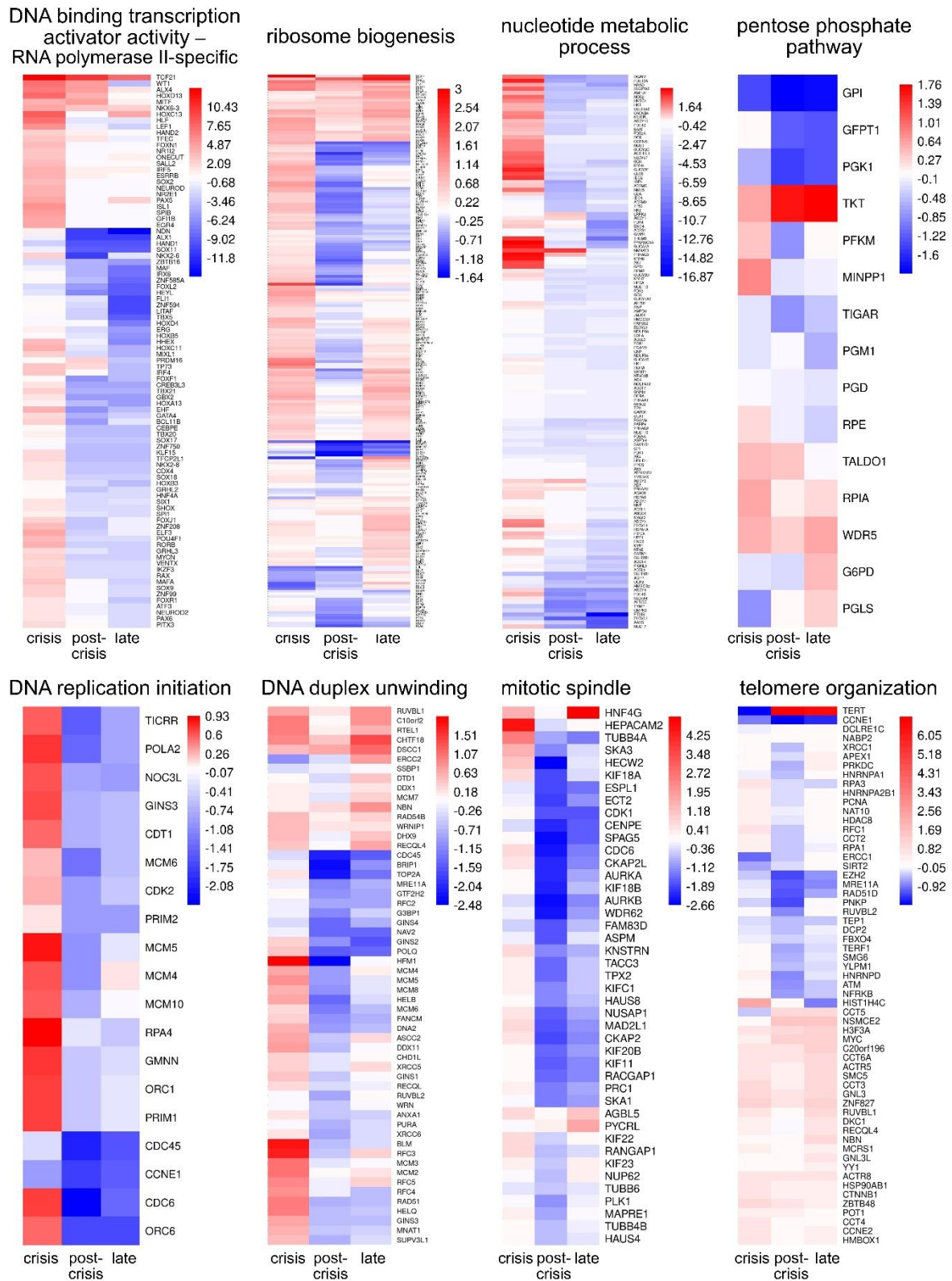


Figure 38. Differential gene expression of selected pathways and biological processes based on bulk RNAseq analysis in LFS041. Each heatmap shows DEGs across three passages: early (p.29), post-crisis (p.65), and late (p.346) passages, each compared to early passage (p.19). The colour

gradient represents the level of gene expression changes derived from NOIseq, with red indicating upregulation (in crisis, post-crisis, or late passages) and blue indicating downregulation (in early passage). DNA-binding transcription activator activity (RNA polymerase II-specific) and DNA replication initiation heatmaps were taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I).

Furthermore, I found a gradual increase in the expression of genes involved in telomere organization and maintenance, with the highest expression levels detected in the last passage. Notably, *TERT* showed a significant upregulated expression, in line with the results shown above.

Taken together, transcriptomic and proteomic profiling of gene expression and biological processes provided deeper understanding and valuable insights into the molecular changes occurring during the progression from early passage through crisis to post-crisis and late passages. The observed hypertranscription, impaired nucleotide metabolism, replication stress, mitotic spindle formation, and telomere maintenance collectively contribute to the mechanism of chromothripsis initiation and progression.

3.11. Role of hypertranscription and nucleotide depletion in replication stress in LFS

To further elucidate the mechanism underlying chromothripsis, I collaborated with Wisam Zaatra from Batsheva Kerem's lab, who conducted a series of experiments (data not shown) using the same LFS fibroblasts. To investigate replication stress dynamics, Wisam applied DNA combing, a high-resolution technique that revealed a decrease in replication fork distance, slower replication rates, and an increase in origin firing already at early passages compared to normal fibroblasts, indicating replication stress. This effect was further exacerbated during the crisis. These findings align with my earlier results, where I observed dysregulation in the ORC, replication initiation, and elevated DNA replication stress.

Previous studies have suggested multiple mechanisms for replication stress, including aberrant origin firing, hypertranscription, and nucleotide pool depletion^{79,162,163}. Notably, all three mechanisms were observed in our data. Building on these findings, we hypothesized that replication stress is driven by nucleotide pool insufficiency and hypertranscription. To test this further, Wisam supplemented LFS041 early-passage cells with exogenous nucleosides (A, U, G, and C). Remarkably, by the time of crisis, nucleoside supplementation significantly rescued replication stress, as evidenced by increased replication rates, longer fork distances. Notably, treated cells exhibited reduced genomic instability, with fewer micronuclei, reduced telomere shortening, and decreased DNA damage, as indicated by lower γ H2AX foci levels.

Consistent with the proteomics and RNA-seq results, we speculated that the hypertranscription observed in early and crisis passages was driven by *TP53* loss, leading to increased nucleotide demand and subsequent depletion during DNA replication. Wisam's analyses of LFS cells

Results

showed transcriptional upregulation already at early passages relative to normal fibroblasts, with an even greater increase during the crisis. Importantly, transcriptional inhibition effectively rescued this phenotype, further supporting the hypothesis that nucleotide pool depletion contributes to replication stress and genomic instability in LFS cells.

In summary, the collaborative work supported the hypothesis that *TP53* loss triggers hypertranscription, which increases nucleotide demand and induces replication-related instability in LFS cells, ultimately contributing to chromothripsis.

3.12. Analysis of hypertranscription in cancer using TCGA data

Building upon our identification of hypertranscription in LFS cells associated with *TP53* loss of function, I sought to determine whether the global increase extends to other *TP53*-deficient cancers, reinforcing our findings and providing a broader biological context for the role of *TP53* in transcriptional regulation. Zatzman et al. investigated hypertranscription in 7494 tumours across 31 cancer types from The Cancer Genome Atlas (TCGA) using RNAmP, a computational method designed to directly measure hypertranscription from bulk DNA and RNA sequencing data (See Methods 6.2.17)¹⁶⁴. I used the publicly available TCGA dataset to analyse the same patient cohort, classifying the samples into two groups: (1) tumours with WT *TP53* and *MDM2* (WT *TP53/MDM2*) and (2) tumours with *TP53* loss, *MDM2* gain, or both (*TP53/MDM2*⁺/*TP53*⁻*MDM2*⁺), hereafter referred to as *TP53*⁻/*MDM2*⁺. This classification included both heterozygous and homozygous CNAs, as the low frequency of homozygous *TP53* copy number losses and *MDM2* gains limits the ability to make a robust comparison. *MDM2* gain was included in the analysis due to its functional similarity to *TP53* loss in promoting tumourigenesis^{34,35}. Thus, I analysed a subset of 6,305 tumour samples, comprising 4032 *TP53*⁻/*MDM2*⁺ and 2273 WT *TP53/MDM2* tumours. To assess hypertranscription, I utilized the hypertranscription fold-change (hytx_fch) values from the RNAmP analysis of Zatzman et al.¹⁶⁴. By comparing these values between the two patient groups, I evaluated hypertranscription both globally (across all cancer types) and individually for each tumour type.

The results revealed a significant increase in hypertranscription in the *TP53*⁻/*MDM2*⁺ group relative to the control group in the pan-cancer context (Figure 39A). However, when examining individual tumour types, I excluded 18 out of 31 cancer types from the analysis due to insufficient sample size in either group, which could compromise the statistical robustness. Of the remaining 13 cancer types, 8 exhibited significantly elevated hypertranscription in *TP53*⁻/*MDM2*⁺ tumours compared to their WT *TP53/MDM2* counterparts (Figure 39B-J).

The magnitude of hypertranscription varied by cancer type, with cervical squamous cell carcinoma and endocervical adenocarcinoma showing the highest mean fold change (mean $\pm$ SEM: 3.04 ± 0.2) in the *TP53*⁻/*MDM2*⁺ group, whereas breast invasive carcinoma displayed a

markedly lower hypertranscription fold change (mean $\pm$ SEM: 1.69 ± 0.04) (Figure 39B-C, respectively).

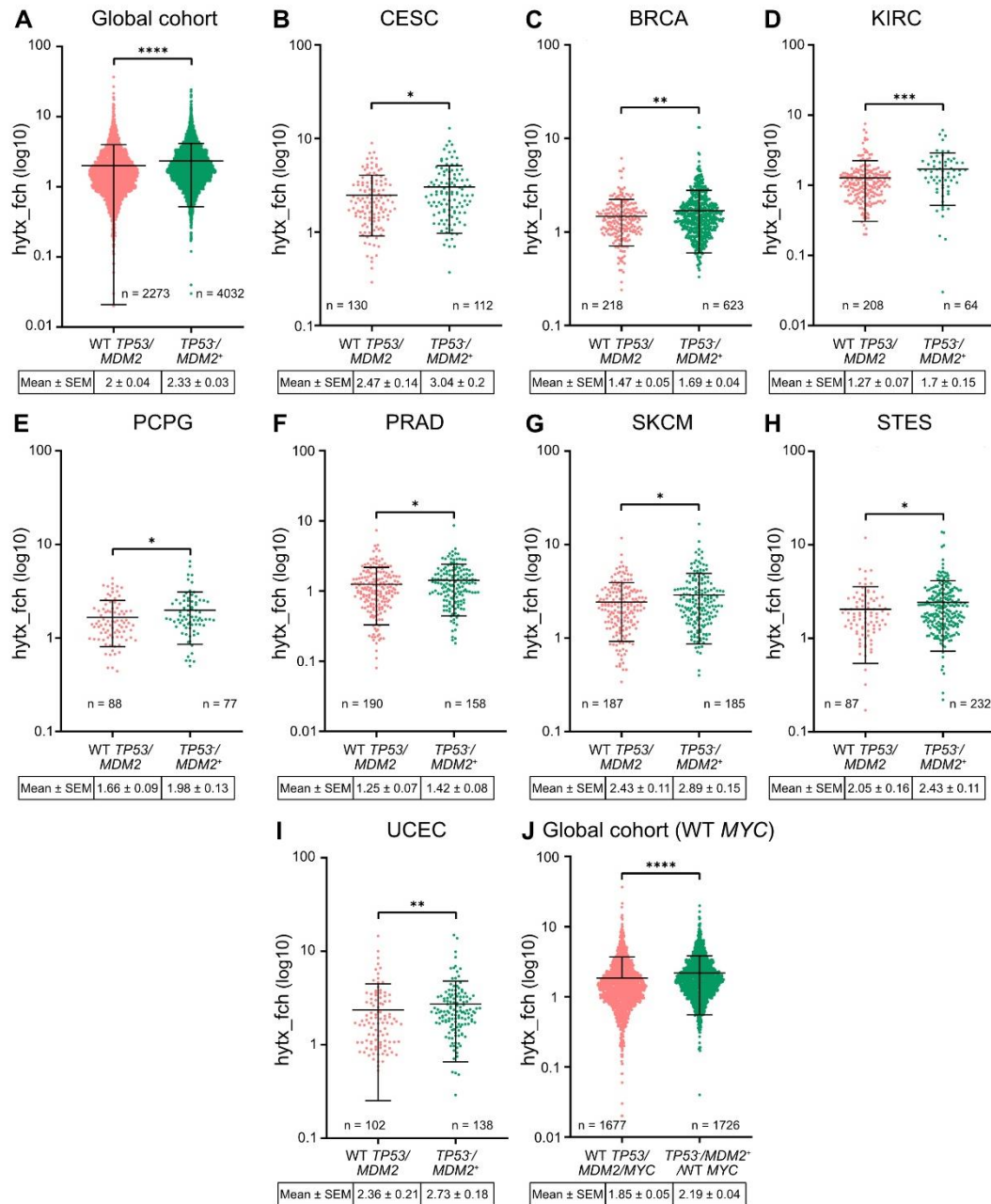


Figure 39. Impact of *TP53/MDM2*⁺ on hypertranscription (reanalysis of TCGA data). **A.** log₁₀-transformed hypertranscription fold-change (hytx_fch) of the TCGA cohort included in the study by Zatzman et al.¹⁶⁴ after I classified them into two groups: WT *TP53/MDM2* and *TP53/MDM2*⁺. Each circle represents a tumour sample, with the total number of samples per group indicated on the plot. **B-I** Cancer-type-specific comparisons of log₁₀ hytx_fch between WT *TP53/MDM2* and *TP53/MDM2*⁺ samples performed for CESC (**B**), BRCA (**C**), KIRC (**D**), PCPG (**E**), PRAD (**F**), SKCM (**G**), STES (**H**), and UCEC (**I**). **J.** hytx_fch analysed in the global cohort under the condition of WT MYC, comparing *TP53/MDM2*⁺/WT MYC to WT *TP53/MDM2*/MYC. Data are presented as mean $\pm$ SEM, with the non-transformed hytx_fch mean value $\pm$ SEM of each group is displayed below the figure. Statistical significance was assessed using a non-parametric t-test (Mann-Whitney test) to compare two independent groups, as the data were not normally distributed. P-values lower than 0.05 were considered statistically significant (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). BRCA: breast invasive carcinoma, CESC: cervical squamous cell carcinoma and endocervical adenocarcinoma, KIRC: kidney renal clear cell carcinoma, PCPG: pheochromocytoma and paraganglioma, PRAD:

Results

prostate adenocarcinoma, SKCM: skin cutaneous melanoma, STES: stomach and oesophageal carcinoma, UCEC: uterine corpus endometrial carcinoma. Figure adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I).

Given that *MYC* is a known driver of hypertranscription^{164,165}, I sought to determine whether the observed hypertranscription was directly linked to *TP53*/*MDM2*⁺ independent of *MYC* status or whether *MYC* mediated the observed hypertranscription. To this end, I removed samples with *MYC* alterations from both groups, retaining only *TP53*/*MDM2*⁺ and WT *TP53*/*MDM2* tumours with WT *MYC*. This resulted in a subset of 3,403 patients. Re-analysing the data of this subset showed that hypertranscription remained significantly elevated in the *TP53*/*MDM2*⁺/WT *MYC* group (mean $\pm$ SEM: 2.19 ± 0.04) compared to WT *TP53*/*MDM2*/*MYC* tumours (mean $\pm$ SEM: 1.85 ± 0.04) (Figure 39J).

In summary, by analysing TCGA data, I corroborated and extended our initial findings in LFS cells, demonstrating that hypertranscription is significantly associated with *TP53* loss and *MDM2* gain across multiple tumour types. The independent effect of *MYC* mutation status suggests that *TP53* inactivation alone is sufficient to drive hypertranscription in cancer.

4. Discussion

Over the past decade, extensive research has been conducted to understand chromothripsis and explore its underlying mechanisms. Despite significant progress, several aspects remain unclear. In particular, the precise cellular and molecular events that initiate chromothripsis are still not well understood, including the early factors and pathways leading to this phenomenon. Although several mechanisms have been proposed, the exact timeline of events, ranging from the earliest stages of chromothripsis initiation to its long-term consequences, including selective advantages provided by chromothripsis, clonal evolution and tumour heterogeneity, remain largely unclear.

4.1. LFS skin-derived fibroblasts as a spontaneous model to study chromothripsis

One of the key challenges in studying chromothripsis lies in its inherent complexity and stochastic nature. Current *in vitro* models to study chromothripsis employ a diverse spectrum of approaches, including induction of mitotic defects or telomere crisis, chemical agents, viral infections, and various forms of external triggers such as radiation, to cause chromothripsis and analyse complex genomic rearrangements^{63,64,66,68,74,98,104–114}.

Even though these cellular models have provided valuable insights, they rely on the artificial induction of chromothripsis, which fails to fully recapitulate its spontaneous and complex nature. Moreover, many current *in vitro* systems lack the temporal resolution required to capture the full sequence of molecular and structural events leading to chromothripsis, particularly during the early stages. These limitations hinder the ability to track and analyse chromosomal rearrangements in real time, making it challenging to fully decipher the heterogeneity observed across chromothriptic clones.

On the other hand, *in vivo* models, including genetically engineered mouse and PDX models, have provided valuable insights into chromothripsis^{58,84,96}. However, these models do not fully mirror human genomic contexts or disease phenotypes, often failing to replicate the spontaneous occurrence of chromothripsis seen in human cells. Additionally, these models are limited by factors such as the complexity of tumour microenvironments, tumour heterogeneity, and challenges in real-time tracking of chromosomal events.

While studies of human cancers have significantly advanced our understanding of chromothripsis, revealing its prevalence and clinical impact across various cancers, they characterize fully developed tumours, long after chromothriptic events have occurred^{23,24,33,51,115,116}. The retrospective nature of such studies limits the ability to capture early initiating events and uncover the causative mechanistic factors driving chromothripsis-induced cancers.

Discussion

To understand the mechanisms underlying chromothripsis, I used a cellular model that mimics the spontaneous nature of chromothripsis without any artificial intervention. Skin-derived fibroblasts from LFS patients allowed me to track the events longitudinally and dissect the full sequence of events starting from early passages, that are carry a *TP53* variant¹⁶⁶. Although LFS cells initially still have one WT functional *TP53* allele, the presence of a single mutated *TP53* allele significantly impairs the tumour-suppressive function^{166–168}. Lynch and Milner demonstrated that *TP53* haploinsufficiency (*TP53*^{+/-}) results in a 75% decrease in p53 mRNA and protein levels compared to the cells with two functional p53 alleles (*TP53*^{+/+})¹⁶⁶. This reduction leads to functional impairments, including decreased G1 cell cycle arrest and decreased apoptosis following stress, explaining the elevated cancer risk in LFS patients and the significant genomic instability observed in these fibroblasts^{166–168}.

As the cells progress through culture, they enter a growth crisis, during which they show a complete loss of WT p53 function (“second *TP53* hit”) and experience heightened levels of genomic and phenotypic instability. Notably, previous research has shown that the complete p53 loss of function enables cells to escape apoptosis². In the post-crisis and late passage stages, chromothripsis was clonally detected and the cells underwent spontaneous immortalization. Thus, the LFS model offers a comprehensive system to study the entire process of chromothripsis, from the initial early passages up to the appearance of a dominant clone carrying chromothriptic chromosomes.

4.2. LFS cells exhibit altered morphology as they progress to crisis and post-crisis passages

The ability of rare subsets of fibroblasts to escape crisis and undergo rapid proliferation suggests that crisis serves as a selective bottleneck, where only a subset of cells has acquired the necessary alterations for immortalization. It is important to note that normal fibroblasts do not bypass the crisis but instead undergo senescence upon reaching their Hayflick limit^{169,170}. The inability of LFS172 fibroblasts to overcome the crisis further highlights that not all cells achieve the necessary genomic or epigenetic changes to escape growth arrest. While Bischof et al. first described fibroblast immortalization in 1990¹²⁰, Tsutsui et al. reported that all six independent cultures of LFS087 failed to spontaneously bypass the crisis¹⁷¹. Even after treatment with Aflatoxin B1 (AFB1), an inducer of DNA damage and mutations that enables cells to bypass senescence and promotes genomic instability, only one third of the cultured cells survived and bypassed the crisis¹⁷¹. In LFS cultures, while some cells successfully bypassed the crisis, the majority of the cells underwent apoptosis, suggesting that not all cells can overcome this bottleneck. Thus, given the stochastic nature of crisis escape, increasing the number of cultured plates could improve the likelihood of detecting rare immortalized clones, as the probability of observing such events rises with a larger population of cells.

A prolonged period of growth arrest aligns with the concept of cellular dormancy, where cells temporarily halt proliferation in response to stress or unfavourable conditions¹⁷². These cultures showed slower growth during crisis passages, which took on average several weeks. However, some cultures appeared to enter a quiescent or dormant state during crisis, persisting for several months before eventually resuming rapid proliferation and expansion. Previous research by Julio A. Aguirre-Ghiso described cellular dormancy and how early-stage tumour cells can enter a non-proliferative state before progressing to malignancy¹⁷³. In addition, by knocking out *TP53* in human gastric organoids, Karlsson et al. demonstrated how early genetic alterations can lead to a prolonged period of genomic instability and clonal evolution, representing a form of latency in tumour development¹⁷⁴. In my study, the prolonged crisis phase observed in some fibroblast cultures may reflect a dormant-like state, during which rare surviving cells accumulate changes necessary for escaping growth arrest and resuming proliferation.

The longitudinal tracking of LFS fibroblast cultures revealed distinct growth dynamics and morphological changes across passages, providing key insights into the progressive nature of genomic instability. I showed that while early-passage cells displayed normal proliferation and typical elongated morphology under the brightfield microscope, fluorescent imaging revealed early signs of genomic stress, including chromatin bridges, micronuclei, and occasionally enlarged nuclei. The complete loss of WT *TP53* during the crisis was reflected not only in nuclear abnormalities – such as increased nuclear size, chromosomal bridges, and micronuclei – but also in cellular morphology detected under the brightfield microscope. The heterogeneous cellular appearance seen during the crisis is consistent with previous findings by Bischoff et al., such as enlarged, irregularly shaped cells alongside apoptotic and necrotic cell populations¹²⁰.

In post-crisis and late passages, the accelerated proliferation and altered morphology propose a transition towards a more transformed phenotype. The shift from an elongated early fibroblast-like morphology to more rounded and aggregated cell shapes suggest cytoskeletal alterations, potentially linked to changes in adhesion properties, which is supported by the proteomic and transcriptomic data. Despite their immortalization, hyperploidy, and high SV burden, these cells did not exhibit the same degree of nuclear and cellular instability observed in earlier passages. This relative stability indicates that cells bypassing crisis activate adaptive mechanisms that preserve genomic and morphological integrity, mitigating the stress encountered during crisis and supporting sustained proliferation and survival. Taken together, the phenotypic instability of LFS fibroblasts underscores crisis as a key transition phase.

Importantly, while all four patient cultures showed a similar progression towards immortalization, patient-specific differences were observed, particularly in morphology and

Discussion

crisis timing. These variations suggest that, although LFS fibroblasts share common pathways of genomic instability, the precise trajectory of transformation is influenced by intrinsic genetic and cellular factors. Moreover, the different *TP53* mutations across patients may contribute to these differences. However, further investigation is needed to determine their specific impact. In addition, the heterogeneity in crisis timing, both between different LFS cultures and within subcultures of one given patient, highlights the stochastic and complex nature of chromothripsis.

4.3. LFS cells show early chromosomal bridges, multipolar spindles, and micronuclei, which increase during the crisis

Chromatin bridges^{49,59–66,128}, micronuclei^{65,66,68,70,71,74} and mitotic defects^{21,29,65,66,129} have all been extensively linked to the mechanisms driving chromothripsis initiation. In my study, the gradual increase of both chromatin bridges and micronuclei formation from early to crisis passages comes in line with the increased genomic instability and the loss of the second *TP53* allele during the crisis phase. As cells approach crisis, telomeres gradually shorten and become deprotected and dysfunctional, resulting in end-to-end chromosomal fusions, BFB cycles, dicentric chromosome formation, and ultimately, heightened genomic instability and apoptosis. The role of p53 loss in promoting telomere crisis has been characterized by Maciejowski et al.⁶³. Moreover, previous studies have shown that the loss of the *TERT* gene further accelerates telomere attrition, thereby exacerbating the onset of telomere crisis^{175,176}. This aligns with the observed increase in genomic instability and BFB cycles in my experimental system.

Notably, the relatively high frequency of chromatin bridges that I observed in the early passages compared to post-crisis and late passages was further corroborated by Strand-seq data, which enabled the detection of BFB cycle patterns. While my Strand-seq analysis of early-passage LFS041 cells revealed BFB patterns across nine different chromosomal arms, the reason behind the notably higher frequencies of these events on 5p and 9p remain unclear. This could reflect an early selective pressure even at the pre-crisis stage, potentially driven by the loss of key genes such as *TERT* (on 5p) and *CDKN2A* (on 9p), which are frequently implicated in genomic instability.

However, further investigation is needed to elucidate the underlying causes driving this chromosomal bias. Additionally, expanding Strand-seq analysis to other LFS-derived cultures will be essential to determine whether this phenomenon is unique to LFS041 or represents a more general feature of telomere crisis. In contrast, consistent with the stabilization of telomeres and reduced genomic instability in post-crisis and late passages, these cells exhibited significantly fewer chromatin bridges and micronuclei, suggesting a decrease in cellular stress and a shift towards a more stable, albeit rearranged, genome.

Chromatin bridges form due BFB cycles and chromosome fusions, leading to segregation errors^{21,65}. These errors result in lagging chromosomes, which are frequently sequestered into micronuclei^{21,65}. Hence, I analysed mitotic events through immunofluorescence staining for pH3/alpha tubulin. In both patients, LFS041 and LFS087, the frequency of multipolar spindle formation was significantly increased, particularly during the crisis, which has been shown to drive CIN in previous studies^{177,178}. A study by Silkworth et al. demonstrated that multipolar spindles and centrosome duplication promote CIN by increasing merotelic kinetochore attachments¹⁷⁸. These faulty attachments lead to lagging chromosomes and chromosome missegregation, further driving genomic instability¹⁷⁸.

Interestingly, p53 loss disrupts cell cycle checkpoints, leading to unchecked centrosome amplification and subsequent multipolar spindle formation, ultimately contributing to CIN^{179–181}. My results were further supported by bulk RNA-seq and proteomics data, which highlighted dysregulation in pathways related to the regulation of centrosome cycle, microtubule cytoskeleton organization involved in mitosis, spindle assembly checkpoints, cell cycle checkpoints regulation, and chromosome segregation. Furthermore, scDNA sequencing (Strand-seq and HIPSD-seq) revealed (whole) chromosomal losses and gains in early and crisis passages, providing additional evidence of multipolar spindle formation and chromosome missegregation events. Silkworth et al. also showed that multipolar spindles increase merotelic attachments, where one kinetochore connects to microtubules from two spindle poles, leading to lagging chromosomes and missegregation during cell division¹⁷⁸. However, the frequency of lagging chromosomes and chromatin bridges detected in this analysis was relatively low, likely due to their occurrence during anaphase^{182,183}, the shortest and most rapid phase of mitosis¹⁸⁴, which makes their identification and quantification challenging.

Overall, my data through phenotypic characterization of LFS fibroblasts, supported by bulk RNA-seq and proteomic analysis, show that the loss of *TP53* results in genomic instability and chromothripsis through two distinct mechanisms. The first mechanism involves telomere attrition and the formation of BFB cycles, leading to chromatin bridges, while the other mechanism stems from the impairment of cell cycle checkpoints, causing multipolar spindle formation. Both mechanisms lead to chromosome missegregation and lagging chromosomes, ultimately contributing to micronuclei formation, CIN, chromothripsis, and aneuploidy.

Beyond proving that the LFS model is a suitable spontaneous system for studying the mechanistic basis of chromothripsis, my findings also offer insights into the timing of these events. To uncover the early events initiating chromothripsis and the subsequent clonal evolution, I employed comprehensive multiomics approaches at the DNA, RNA, and protein levels, providing a comprehensive understanding of the underlying mechanisms and clonal evolution of LFS cells.

4.4. High SV diversity already at early passages and later selection of dominant clones

The observed transition from primarily deletions in early-passage cells of LFS041, as detected by Strand-seq, to a predominance of duplications and complex rearrangements in post-crisis and late passages, aligns with the current understanding of cancer genome evolution, where initial instability is followed by more stable, growth-promoting alterations. This sequence of events has been described in a mouse model of pancreatic ductal adenocarcinoma, where p53 inactivation led to early deletions in progenitor cells, followed by amplifications and increasingly complex rearrangements in later stages¹⁸⁵. Interestingly, even in the early passages, a few clonal events were already observed in LFS cells. However, the heterogeneity was notably high, as evidenced by the presence of rare SVs at elevated frequencies.

By manually annotating the various SV types, I constructed a comprehensive matrix of all genomic alterations in these cells. This approach enabled the exclusion of false positive and false negative events, while enabling the detection of small and rare SVs. Despite the absence of data from the crisis passage, phylogenetic analysis and clustering of cells from early, post-crisis, and late passages based on their SV profiles provided key insights into the extent of clonal expansion, evolutionary divergence, and the timing of mutational events in LFS cells.

The high heterogeneity and diversity observed in early-passage cells, where multiple subclones exhibited diverse SV profiles, suggest that these cells were already experiencing significant cellular and genomic stress prior to crisis. These results were further validated by phylogenetic diversity analysis, pairwise distance distributions, and per-cell SV counts. Conversely, phylogenetic analysis showed that post-crisis cells exhibited high clonality and reduced diversity. This transition suggests that these cells underwent an evolutionary bottleneck during the crisis phase, where selective pressures favour the emergence and expansion of specific clones, ultimately leading to greater homogeneity and clonal dominance.

On the other hand, the increased genetic diversity observed in the late passage, as evidenced by the long branch lengths in the phylogenetic tree, shows substantial differences in structural rearrangements between individual cells. This points to an ongoing evolutionary process and continued accumulation of alterations. This observation is consistent with recent findings from our lab, which demonstrated that chromothripsis occurs in distinct tumour subclones, highlighting its role as a heterogeneous and ongoing process in medulloblastoma evolution⁵¹. Moreover, Valle-Inclán et al. demonstrated that chromothripsis is a continuous mutational process, which was detected subclonally in 74% of osteosarcoma tumours, thereby driving clonal diversification and intratumour heterogeneity⁴². However, it is important to consider that long-term cell culturing could also contribute to the accumulation of SVs observed in these cells.

Notably, the identification of rare early-passage cells (1-3%) clustering with post-crisis and late-passage cells, and sharing common SVs with them, indicates that only a minority of cells acquire the necessary alterations to survive the crisis bottleneck and gain a clonal advantage. Furthermore, the mirror pattern observed on different chromosomes in two early-passage cells suggests missegregation errors, further supporting the results on mitotic errors from immunofluorescence analysis.

4.5. Single-cell Strand-seq analysis reveals chromothripsis, complex rearrangements and aneuploidy already in early-passage cells

The resolution of Strand-seq was sufficient to detect complex chromosomal rearrangements, which were then confirmed by bulk WGS for LFS041 post-crisis passage. Although I detected chromothripsis and other complex SVs in around one third of early-passage cells, only two cases on two different cells exhibited complex SV patterns similar to the ones detected clonally in post-crisis and late passages. These findings show that while genomic instability and chromothripsis arise early in cellular progression, most of these events lack selective advantages or did not survive this catastrophic event. Only specific alterations conferring an advantage persist, with just a few cells giving rise to the winning clones observed in post-crisis and late-passage cells. This pattern is consistent with research reporting that most structural rearrangements undergo negative selection¹⁸⁶. The nearly 14% of early-passage cells showing chromothripsis and other complex SVs on chromosome 5 further support the notion that BFB cycles lead to missegregation errors and subsequently to chromosomal pulverization, as hallmarks of BFB cycles were detected on this chromosome.

Beyond structural rearrangements, numerical chromosomal abnormalities also emerged early in cellular progression. Importantly, Strand-seq data revealed aneuploidy across multiple chromosomes in early-passage cells, with polyploidy observed in one cell. Chromosome counting in metaphase spreads further confirmed aneuploidy in numerous cells from both early and crisis passages in LFS041 and LFS087, emphasizing my previous results on multipolar spindle formation and chromosomal missegregation.

Telomere crisis could potentially explain whole genome doubling observed in these passages. A study from Davoli showed that telomere dysfunction leads to tetraploidization in cells lacking p53, as telomere damage blocks Cdk1/CyclinB activity, leading to prolonged G2 phase without proper division, ultimately bypassing mitosis and re-entering S phase¹⁸⁷. On the other hand, the rapid genomic evolution from a near-diploid to a hyperploid state within just a few passages aligns with the concept of chromothripsis as a punctuated evolutionary event in cancer^{21,22}. Furthermore, the predominance of polyploidy in post-crisis and late-passage cells underscores that whole-genome doubling becomes a clonal event following chromothripsis at late stages in these cells.

4.6. HIPSD-seq analysis of CNAs uncovers key factors driving chromothripsis, cellular survival and progression

Performing HIPSD-seq offered two major advantages. First, as a high-throughput whole-genome sequencing method, it enables the analysis of thousands of cells, allowing for the detection of CNAs, rare events, and clonal evolution. Second, crisis-passage cells were not compatible with the requirements of Strand-seq and BrdU incorporation, due to the slow replication and proliferation rates of these cells.

I showed that the results from HIPSD-seq align closely with the Strand-seq data, with a small subset of early- and crisis-passage cells clustering with post-crisis cells on the UMAP, mirroring the SV-based phylogenetic clustering observed in Strand-seq. Moreover, the copy number profiles inferred from HIPSD-seq revealed strong concordance with the Strand-seq data. Notably, the high heterogeneity and genomic diversity observed in early and crisis passages would likely be masked in bulk WGS data, which averages out a substantial fraction of CNAs and results in seemingly balanced regions. This inference is supported by the study of Ernst et al. (2016)⁴⁹, where early passage cells, analysed by bulk WGS, failed to reflect these genomic alterations at the population level. In addition to LFS041, the detection of additional subclones in HIPSD-seq data from the late passage of LFS087 (p.196) further highlights the ongoing evolutionary process of chromothripsis, as discussed earlier. In turn, this heterogeneity explains the diverse ploidy states observed in the metaphase spreads, which average out to a near-diploid state at the bulk level.

The CNAs detected in the early and crisis passages of LFS041 through HIPSD-seq and Strand-seq provide valuable insights into potential drivers of chromothripsis onset and/or cellular survival. A key observation is the loss of 9p, where the tumour suppressor gene *CDKN2A* (Cyclin Dependent Kinase Inhibitor 2A) is located, which is particularly significant due to its involvement in regulating the cell cycle, apoptosis, and tumour growth, as well as in promoting anti-tumour immunity^{188–190}. Moreover, previous studies also highlighted that deletions involving *CDKN2A* and type I interferons (IFN-I) can lead to immune suppression in certain cancers^{191,192}.

In addition, focal gains observed in post-crisis cells point to genes that may confer a survival and/or clonal advantage to chromothriptic cells. Among those are the gains on 8q, where the oncogene *MYC* is located, as well as at 5p, which contains the *TERT* and *CLPTM1L* genes. These loci have been shown to be amplified in a variety of cancer types¹⁹³. In particular, the *TERT-CLPTM1L* gene locus was found to be amplified in 10 different malignancies in GWAS (genome wide association studies)^{194–196}. Furthermore, *CLPTM1L* overexpression has been linked to promoting aneuploidy, cell proliferation, and survival^{195,197}.

Importantly, the clonal gain of the *TERT* locus in post-crisis cells, even though one copy was lost at earlier passages, suggests that BFB cycles on chromosome 5p might have led to chromothripsis, followed by stabilization via *TERT* gain. In summary, my single-cell resolution analyses enabled a comprehensive, in-depth longitudinal characterization of four distinct time points, representing a spectrum from early to late stages of chromothripsis initiation and progression.

4.7. Bulk WGS identifies recurrent chromosomes affected by chromothripsis and genetic drivers in LFS fibroblasts

Through ShatterSeek analysis of bulk WGS on LFS041 p.63 (post-crisis) and LFS087 p.195 (late), I identified multiple chromothriptic chromosomes in each sample. Notably, chromosome 15 was affected in both patients. The relevance of chromosome 15 remains to be fully explored. While the involvement of chromosome 15 in chromothripsis has been documented in various cancer contexts^{19,198}, it has not been specifically described as enriched for this phenomenon. Nevertheless, the susceptibility to BFB cycles and reported telomere deficiencies of chromosome 15 may contribute to its vulnerability to chromothripsis events^{107,199}.

To gain further insights, it is essential to analyse the other two cultures, LFS172 and LFS174, to determine whether chromothripsis similarly affects chromosome 15 or whether other chromosomes are chromothriptic in LFS041 or LFS087. Additionally, further investigation is needed to understand why these specific chromosomes are recurrently involved. Potential contributing factors may include their association with fragile sites, telomere dysfunction, selective advantages conferred by chromothripsis, or shared intrinsic characteristics of the LFS fibroblasts.

Given that chromothripsis leads to numerous DNA DSBs, the high number of genes affected by breakpoints in each sample (653 genes in LFS041 and 727 genes in LFS087) is consistent with the large-scale genomic rearrangements associated with this phenomenon. The 58 shared genes between the two patients suggest potential recurrent targets of chromothripsis in LFS cells, including both tumour suppressors and oncogenes. Therefore, further investigation is required to elucidate the roles of these genes in chromothripsis-driven oncogenesis, as well as the contributions of the detected drivers identified through OncoAnalyser.

4.8. Telomere stabilization occurs close to the crisis and is essential for cell survival

My analysis of telomere dynamics underscores the crucial role of telomere stabilization in the survival of chromothriptic cells, allowing them to evade senescence and apoptosis to ultimately bypass the crisis bottleneck.

Discussion

The results indicate that a subset of cells in early passages had already undergone *TERT* (LFS041 and LFS174) or ALT pathway (LFS087 and LFS172) activation, which generally increased slightly during crisis. This observation is particularly relevant, as rare ALT-positive events might remain undetected using conventional methods such as the C-circle assay. The clonal *TERT* gain in all LFS041 post-crisis cells and in all LFS174 late-passage cells indicates that *TERT* activation was the main mechanism in both cultures. *TERT* results in LFS041 were corroborated across multiple layers of analyses, including single-cell and bulk WGS, as well as transcriptomic data from bulk RNA-seq. Although the majority of late-passage LFS041 cells exhibited *TERT* gain, the heterogeneity observed in these cells, with some displaying diploid or loss of *TERT* copy number, aligns with the progressive evolution and ongoing genomic instability seen at this stage. On the other hand, even though LFS087 and LFS172 were ALT-positive in the post-crisis and late passages, the substantial ALT-negative fraction in these cells suggests the involvement of other mechanisms within this cellular population. I proved through *TERT*-FISH analysis in LFS172 that nearly 40% of cells exhibited moderate *TERT* gain, with no more than five copies, which would have been undetectable by low-sensitivity methods. It has been previously shown that both ALT and *TERT* mechanisms can coexist not only in the same population but even within individual cells^{200,201}.

It is important to note that ALT pathway activation may vary depending on the cell cycle phase²⁰², which could contribute to its variable detectability. Given that the co-existence of *TERT* and ALT activation within the same population was analysed in only one sample (LFS172 p.72), future studies should extend these investigations to include both *TERT* and ALT analyses across all four patients. This variability in telomere maintenance strategies may represent distinct survival mechanisms or indicate the presence of unique cellular subpopulations with specific molecular characteristics. Clinically, these differences could have significant implications, particularly in relation to cellular resistance to therapies, emphasizing the need for further investigation into telomere maintenance mechanisms in chromothriptic cells. Altogether, telomere stabilization achieved in chromothriptic cells is a critical feature in dominant clones and likely accounts for the low numbers of BFB cycles and chromatin bridges observed in post-crisis and late passages.

4.9. Telomerase activation prevents crisis and reduces chromosomal abnormalities

hTERT transduction in early-passage cells led to enhanced proliferation and a marked reduction in chromosomal abnormalities, including chromatin bridges and micronuclei. Unlike the control cells, the absence of crisis in *hTERT*⁺ cells suggests that telomerase activation effectively prevents replicative senescence, mitigating cellular stress that could otherwise drive genomic instability in LFS cells. The stark contrast in morphology, growth dynamics, and

genomic stability between *hTERT*⁺ and control cells underscores the stabilizing and protective role of telomere maintenance in preserving chromosomal integrity.

Single-cell sequencing on LFS174 p.31 of *hTERT*⁺ and pMX-GFP control cells further confirmed that early telomerase activation significantly reduced genomic alterations, with *hTERT*⁺ cells exhibiting fewer CNAs and less aneuploidy as compared to control cells. While both *hTERT*⁺ and control cells shared a subset of clonal CNAs, these likely arose before transduction, as the cells had already undergone 18 passages by the time of transduction. This is consistent with findings from Ernst et al., who detected similar clonal CNAs in bulk WGS sequencing of p.19 cells (data not shown)⁴⁹. Surprisingly, the clonal 8p loss observed exclusively in control cells matches the loss detected in late-passage LFS174 chromothriptic cells (p.84) in the Ernst et al. study⁴⁹, suggesting that this alteration could be involved in driving genomic instability, potentially linked to chromothripsis events seen in untransduced LFS174 cells.

Although *hTERT* appears to promote a more stable genomic landscape, the resolution of the scDNA sequencing data generated using PTA was insufficient to confidently identify chromothripsis events. The presence of subclonal alterations in both populations highlights the inherent genomic heterogeneity in LFS cells, even with telomerase activation. These findings warrant further investigation, including verifying the p53 status to determine whether the second allele was lost in *hTERT*-transduced cells.

4.10. ecDNAs and gene fusions confer chromothriptic cells selective advantages

Interestingly, I detected ecDNAs in a small subset of early-passage cells in both LFS041 and LFS087. However, the peak in ecDNA-positive cells during crisis, followed by a decline in later passages, suggests that ecDNA formation confers the cells a survival and proliferation advantage to overcome crisis. This pattern aligns with previous studies reporting that ecDNA can drive tumour evolution and contribute to genomic instability²⁰³. The detection of *MYC*-containing ecDNA in post-crisis LFS041 cells demonstrates that ecDNA-mediated *MYC* amplification may have contributed to the selective advantages in the dominant clones. The subsequent decline in ecDNA prevalence in the very late passages could be attributed to ecDNA reintegration into linear chromosomes, which has been previously described^{88,90,204,205}. Additionally, the reduction in ecDNA levels might reflect a cellular response to mitigate the potential burden of excessive ecDNA²⁰⁶.

As expected, chromothripsis resulted in a significant increase in gene fusions in post-crisis and late passages compared to earlier time points, as shown by bulk RNA-seq data. Several notable fusions emerged, including those involving *TRIM40* in the late passage of LFS087. In normal tissues, *TRIM40* expression is barely detectable, whereas its overexpression has been

Discussion

reported to disrupt cell-cell adhesion and act as a pathogenic driver in epithelial cells²⁰⁷. Additionally, *TRIM40* overexpression has also been implicated in gastric²⁰⁸ and oesophagus cancer²⁰⁹. Another notable gene fusion detected in LFS087 late passage involves *ERC1*. *ERC1* fusions have been previously reported in lung cancer, thyroid carcinoma, and pancreatic ductal adenocarcinoma, with *ERC1* also being associated with malignant cellular motility and invasiveness^{210,211}.

Similarly, post-crisis and late-passage cells of LFS041 also exhibited gene fusions, including those involving *COL5A1* gene. *COL5A1* has been shown to drive various oncogenic processes, such as proliferation and invasion, across multiple cancer types, including clear cell renal cell carcinoma²¹², lung adenocarcinoma²¹³, gliomas^{214,215}, ovarian cancer²¹⁶, breast invasive ductal carcinoma²¹⁷ and cervical cancer²¹⁸. These results highlight the role of ecDNAs and gene fusions in generating cancer drivers and in promoting the progression and proliferation of chromothriptic cells.

4.11. Proteomics and transcriptomics data reveal that hypertranscription and metabolic dysregulation result in impaired nucleotide pools and subsequent replication stress

In this study, I employed a multi-omics approach, combining mass spectrometry-based proteomics on LFS041 cells and bulk RNA-seq on both LFS041 and LFS087 cultures. This integrative strategy provided a comprehensive view of the molecular changes occurring during the crisis and post-crisis stages. The concordance between proteomics and transcriptomics data reinforced my findings, while each approach offered unique insights into the involved cellular processes.

Notably, I observed a striking upregulation of genes and proteins in the LFS041 crisis passage compared to the early passage, both at the RNA and protein levels. This widespread increase in expression hinted to a state of hypertranscription during the crisis phase. Interestingly, while the number of significantly upregulated proteins in the late passage of LFS041 was nearly double as compared to the crisis passage, this pattern was not mirrored at the RNA level. This discrepancy suggests a more global transcriptional increase during crisis compared to later passages, potentially indicating post-transcriptional regulation mechanisms at work in the late passage.

Gene set enrichment analysis revealed a significant upregulation of transcription- and translation-related processes in crisis-passage cells compared to early-passage cells. These processes showed a progressive increase from early to crisis to later passages, encompassing various aspects of gene expression. Specifically, I identified significant upregulation in DNA-templated transcription initiation, rRNA metabolic processes, nucleoplasmic transport,

translational initiation, and rRNA processing. Additionally, key regulators of RNA transport and processing, such as *NCL*, *SSB*, *HNRNPD*, *HNRNPA3*, and *RNGTT*, showed increased expression. These findings were further supported by TFEA analysis, which identified *PA2G4* (EBP1 protein) as the top enriched regulator across all comparisons, which is implicated in rRNA processing, ribosome biosynthesis, and transcription^{141–145}.

Collectively, the observed upregulation of transcriptional activity following the loss of *TP53* around the time of crisis strongly suggests that *TP53* loss is a key driving factor of hypertranscription. As a proof-of-concept, I explored the relationship between p53 dysfunction and hypertranscription in cancer by analysing hypertranscription fold change in TCGA data stratified by *TP53* status. Hence, *TP53* loss was consistently associated with increased hypertranscription across all cancer types, as well as within several individual tumour types.

The upregulation in rRNA processing is particularly important, given that rRNA constitutes approximately 85% of the total nucleic acid content in cells, while mRNA accounts for less than 5%²¹⁹. This underscores the significant role of transcription, alongside DNA replication, as a major consumer of nucleotides. The crisis phase exhibited a marked enrichment in nucleotide and nucleoside metabolism and regulation, not only in comparison to early-passage cells but also to post-crisis and late passages. This suggests that the global hypertranscription observed during the crisis phase leads to an increased demand for nucleotide synthesis and consumption, ultimately resulting in nucleotide pool deficiency. This is particularly significant, as Bester et al. from Batsheva Kerem's lab demonstrated that nucleotide depletion contributes to increased genomic instability in early tumour progression²²⁰.

This depletion in nucleotides, in turn, induces replication stress during the crisis phase. Evidence for this is seen in the upregulation of several essential processes involved in DNA replication in crisis-passage cells, which lack functional p53 at this stage. Specifically, the upregulation of processes such as replication initiation, DNA duplex unwinding, and replication fork activity are indicative of replication stress. Furthermore, the enrichment of DNA repair processes in crisis cells compared to early-passage cells aligns with the increased genomic stress experienced during this phase. Previous research found that replication stress can be attributed to transcription-replication conflicts, which occur either when transcription and replication proceed in opposite directions (head-on collisions) or when they move in the same direction but still interfere (co-directional conflicts)²²¹. However, our results suggest that hypertranscription and nucleotide pool dysregulation are causes of p53-induced replication stress.

In addition, the crisis phase was characterized by dysregulation of numerous metabolic processes, including the pentose phosphate pathway, which plays a crucial role in nucleotide synthesis through the generation of ribose 5-phosphate¹³⁹. In addition, *PRMT3* was identified

Discussion

as one of the top three enriched transcriptional regulators through TFEA analysis. This further supports the dysregulation of the pentose phosphate pathway as a cause of decreased nucleotide production, exacerbating the insufficiency in the nucleotide pool and replication stress. Importantly, Mannherz and Agarwal demonstrated that thymidine metabolism plays a critical role in regulating human telomerase activity and telomere length²²². This comes in line with my results, as impaired nucleotide metabolic regulation can lead to telomere shortening, which in turn promotes the formation of BFB cycles and chromatin bridges.

While post-crisis and late passage cells also exhibit hypertranscription, they did not show increased replication stress (as confirmed by our collaborators in the Kerem's lab) or dysregulation in nucleotide metabolism, even when compared to early passage cells. This adaptation could be explained by *MYC* amplification, which directly regulates and enhances nucleotide production²¹⁹, enabling cells to mitigate replication stress.

In addition, p53 loss and the subsequent transcriptional misregulation at the crisis, results in upregulation of peptide antigen processing, interferon signalling and receptor signalling pathway via JAK-STAT, indicating enhanced immune responses and inflammatory response, consistent with previous research²²³. As expected, as a consequence of chromothripsis and rearrangements, DNA conformational changes, chromatin organization, telomere lengthening and post-transcriptional gene regulation were enriched in post-crisis and late passages.

Altogether, my work, supported by results on replication stress, hypertranscription, and nucleotide dysregulation from my collaborator partners, demonstrate that even one functional copy of p53 is insufficient to prevent replication stress and genomic instability. This is evidenced by the genomic stress observed in early-passage cells, which are similar to pre-cancerous lesions. This observation is in line with previous research linking replication damage and stress to the earliest stages of cancer development²²⁴.

Thus, we propose a model, where the loss of the second *TP53* allele triggers a cascade of events, including impaired nucleotide regulation due to hypertranscription, which increases nucleotide consumption and creates a conflict between transcription and DNA replication. This, combined with dysregulation in the pentose phosphate pathway, leads to a reduction in nucleotide production (Figure 40). The depletion of the nucleotide pool during crisis induces DNA replication stress. As a consequence of replication stress and checkpoint dysregulation due to *TP53* loss, cells experience increased telomere attrition, higher DSBs and DNA damage, formation of BFB cycles, dicentric chromosomes through telomere fusion, chromatin bridges, multipolar spindle formation, missegregation errors, and micronuclei formation. These events ultimately lead to chromothripsis.

While most cells undergo apoptosis or negative selection, a small subset of cells survive due to telomere stabilization, ecDNA formation, gene fusions, oncogene activation, and loss of tumour suppressor genes. These adaptations confer survival, selection, and proliferation advantages to chromothriptic cells, driving clonal evolution, expansion, and dominance. Ultimately, these events are suspected to lead to tumourigenesis. Thus, although chromosome shattering occurs as a single catastrophic event, the process of chromothripsis unfolds over a short timeframe and spans only a few population doublings.

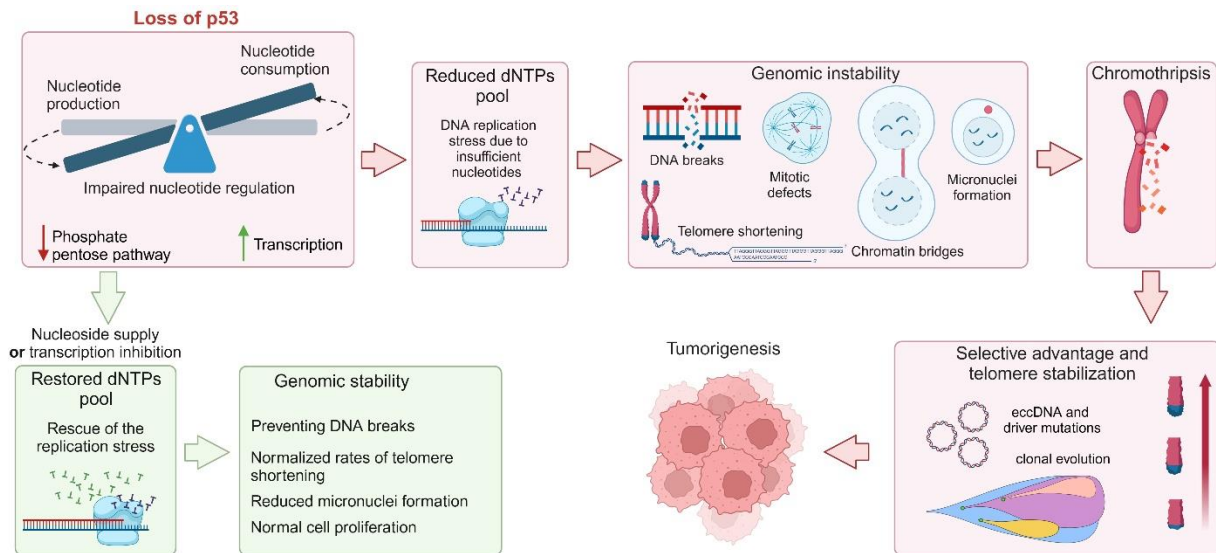


Figure 40. Graphical abstract summarising the main findings. Figure adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Original figure was created with Wisam Zaatra using BioRender.com.

Rescue experiments conducted by my collaborator from the Kerem lab (Wisam Zaatra) provide evidence for the role of nucleotide insufficiency. Their results demonstrate that nucleotide supplementation or inhibition of transcription prevents replication stress, normalizes the rate of telomere shortening, and reduces genomic instability. These findings lay the basis to the development of novel therapeutic interventions targeting these mechanisms.

4.12. Limitations and future perspectives

While LFS cells provide a valuable spontaneous *in vitro* model for studying chromothripsis and genomic instability, they lack the complexity of a tumour microenvironment, including immune infiltration and stromal interactions. Furthermore, these fibroblast-derived cells do not fully recapitulate the morphology of bona fide fibrosarcoma cells, restricting direct clinical extrapolation. Bischoff et al. reported that the skin biopsies were collected years after therapy and from non-irradiated regions, and initial karyotypes and cellular growth appeared normal¹²⁰. However, the possibility of SVs or mutations occurring in a small fraction of cells cannot be excluded¹²⁰.

Further investigation using patient-derived fibroblasts could clarify the long-term impact of chromothripsis on cellular fitness, variations in DNA repair capacity, and telomere dynamics

Discussion

during early stages and post-crisis stabilization mechanisms. One particularly important aspect is nuclear architecture and chromatin state, which may influence fragmentation patterns and contribute to cancer-type-specific heterogeneity.

Assessing the tumourigenicity of LFS fibroblasts is crucial for developing *in vivo* models, but this remains a challenge. While LFS fibroblasts can exhibit anchorage-independent growth *in vitro*, they often fail to form tumours in nude mice¹²⁰. However, Gollahon et al. demonstrated that *RAS*-transformed LFS fibroblasts successfully formed tumours¹²¹, suggesting that specific oncogenic drivers may be necessary for tumourigenicity. Notably, *in vitro* conditions of LFS-derived fibroblasts may impose selective pressures that favour certain cell populations and do not fully reflect *in vivo* tumourigenesis. This raises important questions about the genetic alterations needed not only for chromothriptic cell survival and proliferation, but also for malignant transformation. Identifying the oncogenic drivers and cooperating mutations required for LFS fibroblasts to overcome barriers to tumourigenicity will enhance our understanding of malignant transformation and improve *in vivo* models to study chromothripsis. To bridge remaining knowledge gaps, long-read sequencing could help resolve complex genomic rearrangements, while advanced live-cell imaging may enable real-time tracking of chromothripsis events.

The discovery of the mechanistic role of hypertranscription and nucleotide insufficiency in p53-driven chromothripsis offers novel translational opportunities. Nucleotide supplementation or hypertranscription inhibition could help mitigate replication stress and genomic instability, potentially limiting cancer progression. Combining these strategies with immune checkpoint inhibitors or PARP inhibitors targeting DNA repair deficiencies could potentially provide a multi-faceted therapeutic approach for patients with chromothripsis-driven malignancies.

5. Conclusion

In my study, I used an *in vitro* system of LFS skin-derived fibroblasts with germline p53 variants to investigate genomic instability and spontaneous chromothripsis, without artificial induction. Using a combination of phenotypic, functional, and cutting-edge multiomics approaches, I directly captured each step from the earliest stages of chromothripsis initiation up to the appearance of dominant clones carrying chromothriptic chromosomes.

This work reveals the complex interplay between p53 function, nucleotide metabolism, and genomic stability in the initial phases of cancer development. The data identify hypertranscription as a consequence of WT p53 loss and demonstrate how impaired nucleotide metabolism induces replication stress and initiates a series of events including telomere dysfunction, mitotic defects, ultimately causing chromothripsis. Furthermore, I identified critical features and potential drivers of telomere stabilization, ecDNAs, and gene fusions, which drive the selection and evolution of dominant clones.

In summary, these findings contribute to a deeper understanding of the mechanisms underlying genomic instability in cancer and may potentially guide future approaches to cancer prevention and treatment.

6. Materials and methods

6.1. Materials

6.1.1. Cell lines

Table 1. LFS skin-derived fibroblasts used in the project.

Cell line	Donor	Age (years)	Sex	Cell type	Supplier
LFS041	LFS patient	22	female	skin-derived fibroblasts	Michael A. Tainsky (Detroit, USA)
LFS087	LFS patient	24	male	skin-derived fibroblasts	Michael A. Tainsky (Detroit, USA)
LFS172	LFS patient	26	female	skin-derived fibroblasts	Michael A. Tainsky (Detroit, USA)
LFS174	LFS patient	28	female	skin-derived fibroblasts	Michael A. Tainsky (Detroit, USA)

6.1.2. Oligos used for HIPSD-seq

Table 2. Oligonucleotide primers and adaptors. Abbreviations: phos: 5' phosphorylation, *: PTO modification, **: phosphorothioate bond between the last two nucleotides

Oligo name	Oligo sequence	Assay
Read1	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	HIPSD-seq
phos-Read2	5'-[phos]GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	HIPSD-seq
Blocked phos-ME	5'-[Phos]C*T*G*T*C*T*C*T*T*A*T*A*C*A*[23ddC]	HIPSD-seq
phos-Read1	5'-[Phos]TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	HIPSD&R-seq
phos-ME	5'-[Phos]C*T*G*T*C*T*C*T*T*A*T*A*C*A*C*A*T*C*T	HIPSD&R-seq
Read2	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	HIPSD&R-seq
Illumina PE adaptor-1	5'-[Phos]GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	Strand-seq
Illumina PE adaptor-2	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATC**T-3'	Strand-seq
Illumina PCR Primer PE 1.0	5'-AATGATACGCGACCGAGATCTA-CACTCTTTCCCTACACGACGCTCTTCCGATCT-3'	Strand-seq
Custom multiplexing PCR Primer PE 2.0	5'-CAAGCAGAAGACGGCATACGAGATNNNNNNCGGTCTCGGCATTCTGCTGAACCGCTCTTCCGATCT-3'	Strand-seq

6.1.3. R packages

Table 3. List of R packages used for data analysis.

R Package	Version
AnnotationDbi ²²⁵	1.64.1
ape ²²⁶	5.7-1
BiocManager ^{227,228}	1.30.25
BiocParallel ^{229,230}	1.36.0
biomaRt ²³¹	2.58.2
clusterProfiler ²³²	4.8.3
ComplexHeatmap ²³³	2.18.0
devtools ^{234,235}	2.4.5
DOSE ²³⁶	3.27.1
extrafont ²³⁷	0.19
HMMcopy ²³⁸	1.42.0
HTSanalyseR2 ²³⁹	0.99.19
GenomeInfoDb ²⁴⁰	1.38.8
GenomicAlignments ^{241,242}	1.38.2
GenomicFeatures ^{243,244}	1.54.4
GenomicRanges ^{241,245}	1.54.1
ggnewscale ²⁴⁶	0.5.0
ggplot2 ²⁴⁷	3.5.1
ggpubr ^{248,249}	0.6.0
KEGGREST ^{250,251}	1.42.0
NOISeq ^{153,154}	2.46.0
org.Hs.eg.db ²⁵²	3.18.0
pathview ²⁵³	1.42.0
pheatmap ²⁵⁴	1.0.12
ReactomePA ²⁵⁵	1.46.0
readxl ^{256,257}	1.4.3
ReConPlot ²⁵⁸	1.1
ShatterSeek ²³	0.2

6.1.4. Python packages

Table 4. List of Python packages used for data analysis.

Python Package	Version
pandas ²⁵⁹	2.2.3
Matplotlib ²⁶⁰	3.9.3
NumPy ²⁶¹	2.0.2
Scanpy ²⁶²	1.10.4
SciPy ²⁶³	1.14.1
seaborn ²⁶⁴	0.13.2

6.1.5. Softwares

Table 5. List of software tools.

Software	Supplier/ Company
Affinity Designer 2 (Version 2.5.5)	Serif
AmpliconSuite (v1.2.2)	UC San Diego / AmpliconSuite Team
BioRender	BioRender Inc.
Cell Ranger DNA pipeline (version 1.1.0)	10x Genomics (accessed via DKFZ ODCF Cluster)
Cell Ranger ATAC pipeline (version 2.1.0)	10x Genomics (accessed via DKFZ ODCF Cluster)
Cytoscape (version 3.10.2)	Cytoscape Consortium
FACSDiva 9.8 software	BD Biosciences
GraphPad Prism 8	GraphPad Software
ImageJ 1.54f (Fiji)	National Institutes of Health (NIH)
Leica LAS X Software (v3.7.6)	Leica Microsystems
MaxQuant software (version 2.0.3.0)	Max Planck Institute of Biochemistry
Microsoft Office for Microsoft 365 (Version 2501)	Microsoft
MosaiCatcher (version 2.0.1)	European Molecular Biology Laboratory (EMBL)
Perseus software (version 1.6.2.1)	Max Planck Institute of Biochemistry
Python (version 3.10.14)	Python Software Foundation (accessed via DKFZ ODCF Cluster)
R Studio (version 4.3.0)	Posit (formerly RStudio) (accessed via DKFZ ODCF Cluster)
ZEN 2.3 Microscopy Software (Blue Edition)	Zeiss
ZEN 3.0 Microscopy Software (Blue Edition)	Zeiss
Zotero (version 7.0.11)	Corporation for Digital Scholarship

6.1.6. Databases and online tools

Table 6. List of databases and online tools.

Database	Website
ChatGPT 4.0	https://chatgpt.com/
ChEA3	https://maayanlab.cloud/chea3/
DeepL	https://www.deepl.com/en/translator
GeneCards	https://www.genecards.org/
g:Profiler	https://biit.cs.ut.ee/gprofiler/
Homo sapiens Uniprot database	https://www.uniprot.org/proteomes/UP000005640
Perplexity	https://www.perplexity.ai/
TCGA	https://portal.gdc.cancer.gov/

6.2. Methods

All methods described below, except for sections 6.2.9, 6.2.10, 6.2.12.5, 6.2.16 and 6.2.19 (including their subsections), were taken from a manuscript in preparation, where I am a shared first author, and have been modified as needed for this thesis. Unless specified otherwise, I was the primary author of these sections.

6.2.1. Cell culture

I used skin-derived fibroblasts from four LFS patients (LFS041, LFS087, LFS172 and LFS174), which were provided by Michael A. Tainsky (Table 1). Cells were cultured in Minimum Essential Medium (MEM) Eagle (Sigma-Aldrich, Cat. No. M5650), supplemented with 10% foetal calf serum (FCS), 1% glutamine, and 1% penicillin/streptomycin. In the early stages of the project, MEM Alpha (Life Technologies, Cat. No. 22571020) with the same supplements was used for cell culture. However, since MEM Alpha did not match the medium used by our collaborators, I conducted a series of experiments – including mass spectrometry-based proteomics, bulk RNA-seq, staining, FISH, and morphological analyses (data not shown) – to ensure that the choice of medium did not affect the cells. These experiments confirmed that no medium-dependent effects were detected.

The cells were maintained in 60 mm (5 cm) cell culture dishes (Falcon, Cat. No. 353004) at 37°C with 5% CO₂ in a Hepa Class 100 incubator (Thermo Fisher Scientific). On average, early-passage cells were split once a week upon reaching 90–95% confluency, while post-crisis and late-passage cells required passaging twice per week. In contrast, crisis-passage cells took several weeks per passage to reach confluency (as described in Results 3.2). The medium was renewed every 3–4 days if the cells were not split. For passaging, the medium was removed, and cells were rinsed once with PBS/EDTA. Trypsinization was performed using 0.25% Trypsin/EDTA for 3–5 min, followed by centrifugation at 300g for 5 min. The supernatant was discarded, and the cell pellet was resuspended in fresh medium and split accordingly.

For cryopreservation, cells were harvested by trypsinization, centrifuged, and resuspended in 1.5 ml medium containing 10% DMSO. Cryovials were placed in Mr. Frosty containers at -80°C overnight for gradual cell freezing before being transferred to liquid nitrogen for long-term storage. For thawing, frozen vials were warmed in a 37°C water bath, then transferred dropwise into a 15 ml Falcon tube containing 5 ml of fresh medium, followed by centrifugation at 200g for 5 min. The supernatant was discarded, and the cell pellet was resuspended in culture medium and plated into a 60 mm cell culture dish.

STR profiling was performed to authenticate the cell lines using a multiplex PCR-based method provided by Multiplexion. The analysis confirmed the uniqueness of the cultures and the absence of cross-contamination with other cell lines. In addition, mycoplasma testing was conducted using a PCR-based assay from Eurofins, which confirmed that the cells were mycoplasma-free. To ensure continued absence of contamination, the cells were routinely tested for mycoplasma approximately every two months using the Mycoplasma PCR Detection Kit (abm, Cat. No. G238), following the manufacturer's instructions.

Materials and methods

6.2.2. Cell morphology imaging and analysis

Cell morphology was imaged using a Zeiss Axio Vert.A1 inverted microscope at 5x magnification under brightfield conditions. ZEN 2.3 Microscopy Software (Blue Edition) was used for data acquisition, while ZEN 3.0 Microscopy Software (Blue Edition) was used for image processing.

6.2.3. Growth curves for LFS fibroblasts

The time required to reach confluency was used as a proxy for growth characteristics across all four LFS fibroblast lines (LFS041, LFS087, LFS172, and LFS174), starting from early passages and continuing until post-crisis, except for LFS172, which did not bypass the crisis and underwent complete cell death. In the growth curve graphs, the splitting ratio used at each passage is indicated for the corresponding passage number. A shorter time to confluency at specific passages suggests increased growth rates, while longer times indicate slower growth, reflecting changes in cell population dynamics. Graphs were generated using Microsoft Excel.

6.2.4. Quantification of micronuclei and chromatin bridges

Cells were seeded on coverslips in 6-well plates and fixed with 4% PFA at pH 6.8 for 15 m. Following fixation, the cells were washed twice with 1X PBS for 5 min, rinsed in double distilled water and then in absolute ethanol. The coverslips were air-dried at room temperature before being mounted on slides (Thermo Fisher Scientific, Cat. No. J1800AMNZ) using DAPI Fluoromount-G Mounting Medium (Southern Biotechnology, Cat. No. 0100-20) and left to solidify in the dark at room temperature for 1 hour.

Cell imaging and quantifications were performed using Zeiss AXIO imager 2. ZEN 2.3 Microscopy Software (Blue Edition) was used for data acquisition, while ZEN 3.0 Microscopy Software (Blue Edition) and ImageJ 1.54f were used for image processing. Micronuclei and chromatin bridges were quantified in a minimum of 750 cells from at least three biological replicates per passage. GraphPad Prism 8 software was used to create graphs, perform statistical tests and calculate p-values. Statistical significance was assessed using a one-way ANOVA followed by Tukey's multiple comparisons test, with p-values below 0.05 considered as statistically significant (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

6.2.5. Immunofluorescence analysis of pH3 and Acetyl-alpha-Tubulin and quantification of mitotic defects

Cells were seeded on coverslips in 6-well plates and allowed to reach 60-70% confluency. Once confluency was achieved, the coverslips were rinsed with cell wash PBS (1X PBS, 1% of 1M MgCl₂) and the cells were fixed with 4% PFA at pH 6.8 for 15 min. Following fixation, the cells were washed twice with 50mM NH₄Cl for 5 min each, and twice with 1X PBS with

rocking. Next, a freshly prepared blocking buffer (1X PBS, 5% normal goat serum, 0.3% Triton X-100) was added to the coverslips and incubated for 1 hour at room temperature. After removing the blocking buffer, a mixture of primary antibodies – phospho-histone H3 (Ser10) (6G3) mouse mAb (Cell Signalling, Cat. No. 9706, Lot#10) diluted at 1:400 and Acetyl- α -Tubulin (Lys40) (D20G3) XP rabbit mAb (Cell Signalling, Cat. No. 5335, Lot #5) diluted at 1:800 – was added to each coverslip in an antibody dilution buffer (1X PBS, 1% BSA, 0.3% Triton X-100). The coverslips were then incubated overnight at 4°C. The next day, the coverslips were washed once with 1X PBS for 45 min and twice for 15 min. A secondary antibody mix consisting of goat anti-mouse and goat anti-rabbit antibodies, both diluted at 1:1000 in antibody dilution buffer, was applied to each coverslip and incubated in the dark for 90 min at room temperature. Afterward, each coverslip was washed three times for 15 minutes with 1X PBS, rinsed in ddH₂O, and then in absolute ethanol. Finally, the coverslips were air-dried at room temperature before being mounted on slides using DAPI Fluoromount-G Mounting Medium and left to solidify in the dark at room temperature for 1 hour.

Cell imaging and quantification were performed using a Zeiss AXIO Imager 2 with ZEN 2.3 Microscopy Software (Blue Edition). Representative images were captured using a Leica SP8 confocal microscope. Maximum intensity projections were applied using Leica LAS X software (v3.7.6), and further image processing was performed with both LAS X software and ImageJ 1.54f. A total of 100 cells per patient and per biological replicate across three independent replicates were quantified. GraphPad Prism 8 software was utilized to create graphs, perform statistical tests, and calculate p-values. Statistical significance was assessed using repeated-measures two-way ANOVA followed by uncorrected Fisher's LSD for multiple comparisons; p-values below 0.05 were considered statistically significant (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

6.2.6. TERT-FISH

50,000 cells in supplemented MEM were seeded on a UV-sterilized slide and incubated at 37°C with 5% CO₂. Once the cells reached 60-80% confluency, the slides were rinsed with cell wash PBS and fixed with 4% PFA at pH 6.8 for 15 min. The slides were then rinsed twice with 1X PBS and dehydrated through a series of ethanol solutions: 70% ethanol for 2 min, 90% ethanol for 2 min, and 100% ethanol for 4 min. Slides were left to air dry at room temperature and subsequently stored at -20°C for short-term storage and at -80°C for long-term storage.

Next, the slides were dehydrated again through an ascending ethanol series (70%, 90%, and 100% ethanol, each for 5 min at room temperature). Samples were digested in a pre-warmed digestion solution (0.01 M HCl, 0.02 mg/ml pepsin) for 10 min at 37°C in a water bath. The samples were then washed in 1X PBS for 5 min and post-fixed with 1% PFA on ice for 5 min. After washing in 1X PBS for 10 min, the slides were dehydrated through an ascending ethanol

Materials and methods

series and air-dried. Meanwhile, the DNA probe was prepared by mixing 10 µl of DNA *TERT* probe (Clone RP11-117B23, Source BioScience), which was labeled with Digoxigenin-11-dUTP (Roche, Cat. No. 1570013), Cot1-DNA (Life technologies, Cat. No. 15279-011), Herring Sperm DNA (Thermo Fisher Scientific, Cat. No. 15634017) and 1/20th volume ratio of 3 M sodium acetate (NaAc) with 2.5X the volume of 100% ethanol. This mixture was precipitated at -80°C for 30 min, centrifuged, and washed with 70% ethanol before air drying at 37°C. 10 µl of deionized formamide were added to the pellet and mixed well by shaking at 1000 rpm for 15 min, followed by adding 10 µl of hybridization mix per slide (20% dextran sulfate in 4X SSC pH 7.0) and shaken for at least 15 min before further processing. After air drying the slides, the probe was applied onto the slides, covered with a coverslip, and sealed with Fixogum. The slides were denatured on a heating block at 78°C for 10 min and incubated overnight in a humidified chamber at 37°C. The next day, the slides were washed with a freshly prepared Wash A solution (50% formamide in 2X SSC pH 7.0) in a water bath at 42°C with shaking for 10 min to remove coverslips, followed by three additional 5-min washes. Next, the slides were washed with Wash B solution (0.5X SSC, pH 7.0, prewarmed at 60°C) at 42°C for three 5-min washes, with gentle shaking. Blocking involved adding blocking buffer (200 µl of 4X SSC containing 3% BSA, pH 7.3) to the slide, covering them with parafilm, and incubating in a humidified chamber at 37°C for 30 min. For detection, 2.8 µl Anti-Digoxigenin-Rhodamine (Roche, Cat. No. 11207750910, 400 µg/ml stock concentration) in 200 µl detection buffer (blocking buffer diluted in 1:3 ratio) were added to the slides, which were then covered with parafilm and incubated in the humidified chamber at 37°C for 30 min, followed by incubation and washing with Wash C solution (0.1% Tween 20 in 4X SSC, pH 7.3) for three 5-min washes at 42°C while shaking in the dark. Finally, slides were mounted using VECTASHIELD Vibrance Antifade Mounting Medium with DAPI (Vector Laboratories, Cat. No. H-1800).

Cell imaging and quantifications were performed using Zeiss AXIO imager 2 and ZEN 2.3 Microscopy Software (Blue Edition) for data acquisition, while ZEN 3.0 Microscopy Software (Blue Edition) and ImageJ 1.54f were used for image processing. 100 cells from each passage were quantified. GraphPad Prism 8 software was used to create graphs, perform statistical tests, and calculate p-values. Statistical significance was assessed using a nonparametric Kruskal-Wallis test, followed by Dunn's multiple comparisons test; p-values less below 0.05 were considered statistically significant (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

6.2.7. ALT FISH

The ALT-FISH protocol was implemented as previously described²⁶⁵, with slight modifications outlined below. A total of 50,000 cells in supplemented MEM were seeded onto UV-sterilized slides and incubated at 37°C with 5% CO₂. Later, slides were washed twice with PBS and then fixed with 70% ice-cold ethanol for 20 min at room temperature. Next, the slides were washed

twice with a washing buffer (100 mM Tris–HCl pH 8, 150 mM NaCl, 0.05% Tween-20), after which the remaining buffer was carefully removed.

A hybridization mix containing 2X SSC buffer, 8% deionized formamide, and 5 nM fluorescent DNA probe (5'-Atto594-(CCCTAA)5-3', TelC probe; Eurofins Genomics) was added to the slides, which were incubated for 20 min at 37°C. Following hybridization, the slides were rinsed twice in 2X SSC for 5 min in dark cuvettes. Afterwards, slides were rinsed in ddH₂O, followed by 2 min incubation in 70% ethanol then for 2 min in 100% ethanol in the dark. After the slides were air-dried, they were mounted using VECTASHIELD Vibrance Antifade Mounting Medium with DAPI.

Quantification was performed on 100 cells per patient and per biological replicate across two independent replicates using a Zeiss AXIO Imager 2 microscope and ZEN 2.3 Microscopy Software (Blue Edition). Representative images were captured using a Leica SP8 confocal microscope. Maximum intensity projections were applied using LAS X software, and further image processing was performed with both LAS X software and ImageJ 1.54f. GraphPad Prism 8 software was used to create graphs, perform statistical tests, and calculate p-values. All values from both biological replicates are shown in the violin plots. Statistical significance was assessed using a nonparametric Kruskal-Wallis test, followed by Dunn's multiple comparisons test; p-values less below 0.05 were considered statistically significant (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

6.2.8. Quantification of ecDNAs and ploidy assessment

Metaphase spreads were prepared by adding 0.04 µg/ml colcemid to dividing cells (~60% confluent) and cells were incubated for 8 hours at 37°C. Next, the supernatant was collected in a 15 ml Falcon tube, centrifuged and resuspended in 500 µl. 10-15 ml of a hypotonic solution (KCL 0.4% – prewarmed in a water bath) was added in a dropwise manner while gently mixing and incubated for 25 min at 37°C. After centrifugation at 300g for 10 min, supernatant was removed, leaving 2-3 ml to resuspend the cell pellet. Cells were then fixated by adding 7-8 ml of ice-cold fixative solution (Methanol 3:1 glacial acetic acid) in a dropwise manner while gently mixing, followed by centrifugation at 300g for 10 min and discarding the supernatant. Fixation steps were repeated twice. After discarding the supernatant, cells were resuspended in 200-500 µl of the fixative solution, depending on the size of the pellet. For slide preparation, 20 µl of the cell suspension were dropped onto humidified pre-cleaned slides and allowed to air-dry. Slides were stored at -20°C for short-term storage and at -80°C for long-term storage. VECTASHIELD Vibrance Antifade Mounting Medium with DAPI was used for mounting the slides.

Materials and methods

Ploidy and ecDNA quantifications were performed using Zeiss AXIO imager 2 and ZEN 2.3 Microscopy Software (Blue Edition). For ploidy counts, images were captured and analysed using graphic tools on ZEN 2.3 Microscopy Software (Blue Edition) to identify the number of chromosomes per cell. 50 metaphases were quantified for ecDNAs and ploidy. GraphPad Prism 8 software was used to create graphs for ecDNAs, while Microsoft Excel was used to create ploidy graphs. Representative images of ecDNAs were acquired using Leica SP8 confocal microscope and data acquisition was performed using LAS X software. ImageJ 1.54f was used for further image processing.

6.2.9. Retroviral transduction protocol of *hTERT*

The original protocol was provided by Florian Selt and adapted by me. Retroviral transduction was performed on p.18 (early) of all four LFS lines (LFS041, LFS087, LFS172 and LFS174). I performed growth curve analysis on all cell lines using a range of hygromycin concentrations (1000, 500, 250, 125, 62.5, 31.25 µg/ml) to determine the lowest effective concentration for selecting transduced with minimal cellular toxicity.

Retroviral transduction was performed as follows: On day 0, Platinum GP packaging cells were seeded at a density of 7×10^6 cells per 10 cm cell culture dish, without adding Penicillin/Streptomycin, Puromycin, or Blasticidin. On day 1, A-Mix was prepared by adding diluting 5 µg of retroviral vector DNA (pBabe-hygro *empty* (Addgene, Cat. No. 1765), pBabe_hygro *hTERT* (Addgene, Cat. No. 1773), or pMX-GFP (CellBioLabs, Cat. No. RTV-050) in 500 µl of serum-free DMEM medium, together with 2.5 µg of VSV-G envelope vector (CellBioLabs, Cat. No. RV-110), while B-Mix was prepared by adding 22,5 µg of 1X PEI (Polyethylenimine; 1mg/ml) to 500µl serum-free DMEM. The two mixes were combined, mixed thoroughly, and incubated at room temperature for 20 min. The transfection mixture was gradually added in a dropwise manner to the Platinum GP cells. The pMX-GFP vector was included as an additional transfection control, as no fluorescent marker was present on pBabe vectors (*empty* and *hTERT*). The following day (day 2), the culture medium of Platinum GP cells was replaced with fresh DMEM supplemented with 10% FCS. For each virus seed 5×10^5 LFS cells were seeded in a 5 cm dishes in a supplemented MEM. Mock-infected control cells were also seeded to assess antibiotic selection efficiency. On day 3, the viral supernatant from Platinum GP cells was collected 48 hours after transfection (T1), filtered through a 0.45 µm sterile filter, and supplemented with Polybrene to a final concentration of 8 µg/ml. Fresh DMEM was added to Platinum GP cells to sustain viral production. Viral supernatant that was not used immediately was stored at -80°C. Noting that the virus infectivity is decreased by 50% after each freeze-thaw cycle.

For the first transduction, the medium was removed from LFS cells and replaced with 3 ml of viral supernatant. LFS cells were incubated with viral particles for at least 6 hours, after which supplemented MEM was added, and the cells were incubated overnight. Mock-infected cells were treated under identical conditions, except that MEM with Polybrene was added without the virus. On day 4, 72-hour post-transfection (T2) viral supernatant was collected and filtered as described above. A second round of transduction was performed following the same procedure as on day 3. On day 6, cells were either passaged upon reaching confluence or subjected to antibiotic selection with hygromycin to isolate successfully transduced cells. On the next day (day 7), if the cells had been split on the previous day, hygromycin selection was initiated at an optimal confluency of approximately 25% to promote selective outgrowth of transduced populations.

6.2.10. Cumulative PDLs of transduced cells

Cumulative population doubling levels (PDLs) were calculated for *hTERT*⁺, as well as for transduced control cells with either pMX-GFP or the empty vector. These calculations were performed for LFS041, LFS172, and LFS174. However, in LFS087, both the pMX-GFP and empty vector control cells did not survive. Cell numbers were counted at the time of seeding and upon reaching confluency. PDL was determined using the following equation²⁶⁶:

$$\text{PDL} = 3.32 \times (\log_{10} [\text{final cell count}] - \log_{10} [\text{initial cell count}]) + \text{PDL of the previous passage}$$

The final cell count refers to the number of cells at confluency, and the initial cell count corresponds to the number of cells seeded. The PDL from the previous passage was included to account for cumulative cell divisions across passages. Growth curve graphs were generated using GraphPad Prism 8.

6.2.11. Bulk RNA-sequencing

6.2.11.1. RNA extraction and sequencing

Cell pellets from two LFS patients [LFS041 p.19 (early), LFS041 p.29 (crisis), LFS041 p.65 (post-crisis), LFS041 p.346 (late), LFS087 p.19 (early), LFS087 p.47 (post-crisis) and LFS087 p.195 (late)] were prepared after trypsinization in 0.25% Trypsin/EDTA, cell resuspension and centrifugation. The cell pellets were kept at -80°C until performing RNA extraction using AllPrep DNA/RNA/Protein Mini Kit (QIAGEN, Cat. No. 80004), following the manufacturer's instructions. RNA concentration was measured using NanoDrop, while a Bioanalyser (Agilent RNA 6000 Pico Kit, Cat. No. 5067-1513) was used to measure the quality of the RNA.

Sequencing libraries were prepared with the Illumina TruSeq mRNA stranded Kit following the manufacturer's instructions. In brief, mRNA was purified from 500 ng of total RNA using oligo(dT) beads. Poly(A)⁺ RNA was fragmented to 150 bp and converted to cDNA. The cDNA

Materials and methods

fragments underwent end-repair, 3' adenylation, adapter ligation, and they were subsequently amplified (15 cycles of PCR). The final libraries were validated using the Qubit® RNA BR Assay Kit (Thermo Fisher Scientific, Cat. No. Q10211) and the TapeStation RNA ScreenTape assay (Agilent Technologies, Cat. No. 5067-5576, 5067-5577). 2X 100 bp paired-end sequencing was performed on the Illumina HiSeq 4000 following the manufacturer's protocol. RNA Library preparation and sequencing were performed by the DKFZ NGS core facility. All passages from the same cell line were processed and sequenced together.

6.2.11.2. bulk RNA-seq data processing and alignment

DKFZ/ODCF RNAseq workflow (version 1.3.0) was used to analyse RNA sequencing data²⁶⁷. The specific versions of the components used include the alignment and quality control workflows (version 1.2.73-3)²⁶⁸, Roddy default plugin (version 1.2.2)²⁶⁹, Roddy base plugin (version 1.2.1)²⁷⁰, and Roddy (version 3.5.9)²⁷¹.

In brief, a two-pass alignment approach with the STAR aligner was used to align FASTQ reads for individual samples (version 2.7.10a)²⁷². The alignment was performed to a STAR index created from the 1000 Genomes assembly and gencode version 19 gene models, utilizing sjdbOverhang of 200. Duplicate reads in the main alignment file were marked using Sambamba (version 0.6.5)²⁷³. Samtools was used to sort the Chimeric file (version 1.6)²⁷⁴, followed by duplicate marking and generating of BAM files with Sambamba. Quality control was carried out using the Samtools flagstat command and the RNA-SeQC tool (version 1.1.8)²⁷⁵. Gene-specific read counting over exon features was conducted using Subreads (version 1.6.5)²⁷⁶, based on the Gencode v19 models. A custom script was used to compute RPKM and TPM expression values. Gene fusion detection was performed using Arriba²⁷⁷.

6.2.11.3. Differential gene expression and GSEA analyses

Differential gene expression analysis was performed using the NOISeq package^{153,154} within R studio, in order to compare between passages within individual patients. I used the following key parameters: no normalization (norm = 'n'), a probability threshold for non-replicability (pnr = 0.2), a noise threshold (v = 0.02), and a minimum expression level (lc = 1), to ensure robust detection of significant expression changes.

Gene set enrichment analysis (GSEA) was performed using the clusterProfiler package²³² in R studio to conduct GO, KEGG pathway, and Reactome pathway analyses, with gene annotations supported by org.Hs.eg.db²⁵². Pathways were considered significant if they had an adjusted p-value (FDR) below 0.05, determined by the Kolmogorov–Smirnov test with Benjamini-Hochberg correction. Only pathways meeting this criterion were included for further examination. Pathways with an adjusted p-value greater than 0.05 were excluded to maintain a focus on statistically relevant results. Plots of GSEA results were created using ggplot2²⁴⁷ for

visualizing enriched pathways. Heatmaps illustrating protein expression patterns were generated using the pheatmap R package²⁵⁴, by selecting gene sets based on specific biological processes identified in GSEA-GO results or those hypothesized to play a pivotal role in the mechanisms underlying chromothripsis, even if not significantly enriched in GSEA. Gene expression data were visualized by clustering genes and samples based on their expression patterns using hierarchical clustering with Euclidean distance.

6.2.12. Proteomics analysis of LFS fibroblasts

6.2.12.1. Sample preparation for mass spectrometry analysis

This section was written jointly by me and Gianluca Sigismundo. Cell pellets from LFS fibroblasts [LFS041 p.19 (early), LFS041 p.27 (crisis) and LFS041 p.346 (late)] were trypsinized with 0.25% Trypsin/EDTA, resuspension, and centrifugation. Two replicates were collected from each passage and the cell pellets were stored at -80°C until further processing. Cell pellets were resuspended in 150 µL of 0.1% RapiGest SF Protein Digestion Surfactant (Waters Corporation) in 100mM ammonium bicarbonate (AmBic) in H₂O containing 10mM chloroacetamide (CAA, Sigma) and 40mM tris(2-carboxyethyl)phosphine (TCEP). The suspension was sonicated at 4°C for 15 cycles (30" ON/30" OFF) using a PicoBioruptor (Diagenode). After quantification using a BCA assay (Thermo Fisher Scientific), Samples were heated to 90°C for 5 min and subsequently underwent tryptic digestion (Promega) at 37°C for 18 hours. The pH was adjusted to ~2 using trifluoroacetic acid (TFA), and samples were incubated at 37°C for 30 min before centrifugation at 18,000g for 30 min at 4°C. Next, the supernatants were moved to fresh PCR tubes, and the buffer was replaced with 5 mM triethylammonium bicarbonate (TEAB) using the SP3 protein clean-up protocol^{278,279}.

Peptides were labeled with TMT10plex reagents (Thermo Fisher Scientific) following the manufacturer's instructions: Sample peptides were combined with TMT labeling reagents and incubated at room temperature for 1 hour, subsequently quenched with 5% hydroxylamine for 15 min. The labeled peptides were combined into TMT10plex sets and speedvac was used for air-drying the samples. 100 µl TFA 0.1% was used to resuspend the sample, which were then fractionated under high pH conditions using an Agilent 1200 Infinity HPLC system equipped with a Gemini C18 column (3 µm, 110 Å, 100 x 1.0 mm, Phenomenex). Peptides were separated with a 60-minute linear gradient of 0–35% (v/v) acetonitrile in 20 mM ammonium formate (pH 10) at a flow rate of 0.1 ml/min. Elution of peptides was monitored at 254 nm using a UV detector with a variable wavelength. A total of forty fractions were collected and then pooled into eight fractions.

6.2.12.2. Mass spectrometry data acquisition

This section was primarily written by Gianluca Sigismondo and adapted by me with minor additions. 0.1% trifluoroacetic acid in H₂O was used to resuspend the dried fractions. The resuspended samples were initially loaded onto a trap column (PepMap100 C18 Nano-Trap, dimensions: 100 µm x 2 cm). Subsequently, peptide separation was performed using a 25 cm analytical column (Waters nanoEase BEH C18, dimensions: 75 µm x 250 mm, particle size: 1.7 µm, pore size: 130 Å). The chromatographic separation was carried out using a Thermo Easy nLC 1200 system (Thermo Fisher Scientific) coupled to a nanospray source. The HPLC mobile phase composition consisted of solvent A (water with 0.1% formic acid) and solvent B (80% acetonitrile, 0.1% formic acid). The elution gradient was programmed with a linear increase of solvent B: 3% to 8% over 13 min, 8% to 16% over 21 min, 16% to 50% over 119 min, and 50% to 95% over 10 min. The gradient was held at 95% for 8 min, followed by a decrease to 3% over the final 9 min. A Tri-Hybrid Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) was used to perform peptide analysis, which was operated in positive data-dependent acquisition mode with HCD fragmentation.

Both MS1 and MS2 scans were acquired in the Orbitrap analyser with a 3-second cycle time. MS1 scans were performed at a resolution of 60,000, with an AGC target of 1E6, maximum injection time of 50 ms, and a scan range of 375-1500 m/z. Peptides carrying charge states ranging from 2 to 4 were chosen for fragmentation, with a 60-second exclusion period. MS2 was carried out with a collision energy (CE) of 30%, detected in topN mode, with the first mass set at 110 m/z. The AGC target for MS2 was 2E4, with a maximum injection time of 94 ms.

6.2.12.3. Mass spectrometry data processing analysis and visualization

This section was primarily written by Gianluca Sigismondo and adapted by me with minor additions. RAW data were analysed using MaxQuant software, which includes the Andromeda search engine^{280,281}. Peptide identification was performed by searching against the Homo sapiens Uniprot database, combined with a database of contaminant protein sequences (canonical and isoform). MaxQuant's default parameters were applied, with the following adjustments: Trypsin/P and LysC as digestion enzymes, methionine oxidation and N-terminal acetylation as variable modifications, and cysteine carbamidomethylation as a fixed modification. The Orbitrap instrument settings included a precursor tolerance of 20 ppm and MS tolerance of 0.5 Da. A false discovery rate (FDR) of 1% was applied at both the protein and peptide levels. The match between runs option was enabled, and Label-Free Quantification (LFQ) and iBAQ values were calculated. Subsequent protein analysis was performed using Perseus software²⁸². The dataset was filtered to remove potential contaminants, reverse proteins, and proteins identified only by sites.

Further analysis included only those proteins that were identified by at least one unique peptide in the biological replicates. Intensity values underwent normalization to correct for sample mixing errors. In cases involving multiple TMT experiments, linear modelling-based batch effect correction was applied to account for batch-induced variation. For volcano plot generation, two-sided t-test statistics were employed based on TMT quantitative information of expressed proteins, with an FDR of 0.05 and S0 constant of 0.1. Protein abundance data were transformed into Z-scores, which were calculated for each protein across the samples, and heatmaps were generated to display standardized expression patterns.

6.2.12.4. Downstream proteomic GSEA and Cytoscape analyses

Proteomic data were processed by initially converting gene IDs to Ensembl IDs using g:Profiler²⁸³. In cases where a gene ID was not recognized by g:Profiler, I manually matched it using GeneCards²⁸⁴. GSEA was performed as described above. Heatmaps illustrating protein expression patterns were generated using pheatmap package in R studio²⁵⁴, by selecting gene sets based on specific biological processes identified in GSEA-GO results or those hypothesized to play a pivotal role in the mechanism, even if not significantly enriched in GSEA. Protein expression data were visualized by clustering genes and samples based on their expression patterns using hierarchical clustering with Euclidean distance.

Network-based visualization of enrichment results was conducted using EnrichmentMap app (version 3.5.0)¹³⁷ and AutoAnnotate (version 1.5.1)¹³⁸ in Cytoscape (version 3.10.2)²⁸⁵. The GSEA output files for GO and Reactome analyses were exported from R studio and reformatted into tab-delimited text files to ensure compatibility with input requirements of EnrichmentMap. These files were subsequently analysed in Cytoscape to visualize clustered interactions between enriched biological processes, as well as to uncover their interconnections.

6.2.12.5. Transcription factor enrichment analysis

Transcription factor enrichment analysis (TFEA) was performed on significantly differentially expressed proteins using the ChEA3 (ChIP-X Enrichment Analysis Version 3) web-based tool¹⁴⁰, which utilizes a curated collection of transcription factor-target gene relationships to identify potential upstream regulators. statistical significance was assessed using Fisher's exact test, and results were visualized as network plots.

6.2.13. Bulk whole-genome sequencing (WGS)

6.2.13.1. DNA extraction and sequencing

Cell pellets from LFS fibroblasts (LFS041 p.63, post-crisis and LFS087 p.195, late) were prepared after trypsinization in 0.25% trypsin, cell resuspension and centrifugation. Cell pellets were kept at -80°C until performing DNA extraction using DNeasy Blood & Tissue Kit (QIAGEN,

Materials and methods

Cat. No.: 69504), following manufacturer's instructions. stranded DNA assay Qubit dsDNA HS Assay Kit-100 assays (Life Technologies, Cat. No. Q32851) was used to quantify the DNA, while a Bioanalyser (Agilent High Sensitivity DNA, Cat. No. 5067-4626) was used to measure the quality of the DNA.

Sequencing libraries were prepared using the Illumina TruSeq HT Library Prep Kit following the manufacturer's instructions. Briefly, 100 ng of genomic DNA was fragmented to ~350 bp using a Covaris ultrasonicator (Covaris, Inc.). Following fragmentation, the DNA underwent end-repair, size selection using magnetic beads, extension with an 'A' base to the 3' end, and ligation with TruSeq paired-end indexing adapters. The adapter-ligated libraries were then enriched by 8 PCR cycles and purified 1-2 times using magnetic beads. The generated libraries were validated using the Qubit® dsDNA BR Assay Kit (Thermo Fisher Scientific, Cat. No. Q32853) and using the TapeStation Genomic DNA ScreenTape assay (Agilent Technologies, Cat. No. 5067-5365, 5067-5366). Whole genome sequencing was performed using the Illumina NovaSeq 6K paired-end 150 SP platform. DNA Library preparation and sequencing were carried out by the DKFZ NGS core facility.

6.2.13.2. Analysis of bulk whole-genome sequencing data

Whole-genome sequencing data were processed by the DKFZ OTP pipeline^{286,287}. Briefly, reads were aligned to the 1000 Genome project version of the GRCh37 (hg19) reference genome²⁸⁸ using BWA-MEM (version 0.7.15)²⁸⁹. Picard (version 2.25.1)²⁹⁰ was used to mark duplicates, which were later removed with SAMtools²⁹¹ together with low quality reads. The resultant BAM files were used as input to the analyses described below.

6.2.13.3. OncoAnalyser

This section, along with sections 5.2.13.4 and 5.2.13.5, was written by Petr Smirnov. Bulk WGS derived BAM files were then run through the nf-core OncoAnalyser nextflow pipeline (v0.5.0)²⁹² based on tools developed by the Hartwig Medical Foundation, running in tumour-only mode. In brief, this pipeline calls SNVs using the HMF developed mutation-caller SAGE²⁹³, SVs using GRIDSS²⁹⁴, haplotype resolved copy number variation using PURPLE²⁹⁵, and then clusters and annotates SV clusters using LINX²⁹³.

6.2.13.4. AmpliconSuite

AmpliconSuite²⁹⁶ was run on the BAM files generated by the DKFZ OTP (described above) to identify putative highly amplified extrachromosomal DNA fragments. CNVkit (v0.9.10)²⁹⁷ was first used to segment read counts across the genome and detect copy number changes using the circular binary segmentation algorithm²⁹⁸. Copy number altered segments with copy state of at least 4.5 and minimum length of 50 kb were then used as seeds to AmpliconArchitect²⁹⁹

to reconstruct the structures of these focally amplified regions, followed by AmpliconClassifier to obtain a list of possible circular and ecDNA structures and BFB cycles.

6.2.13.5. Chromothripsis detection

To identify chromothriptic chromosomes, the CNV calls from OncoAnalyser were first filtered for events larger than 500 bases and then used as input the ShatterSeek R package²³ together with the corresponding somatic SVs. This method first identifies the largest SV event cluster by forming an SV-graph where individual events are connected when their breakpoints are interleaved and then selecting the largest connected component per chromosome. The distributions of SV breakpoints and end-joining orientations are then tested against a series of criteria, looking for evidence for or against the random distribution expected in a chromothripsis event. The number of CNV events overlapping the regions affected by this SV cluster is also calculated, and this information is combined to determine a label for chromothripsis in each chromosome, as described in more detail in Cortés-Ciriano et al.²³. For our application of this method, a CNV cutoff of 7 oscillating events (between 2 or 3 states) was used, along with an FDR of 0.2 for determining statistical significance. The OncoAnalyser SVs and CNVs, together with the statistical criteria calculated by ShatterSeek were then visualized using the ReConPlot R package²⁵⁸.

6.2.14. Single-cell DNA sequencing by HIPSD-seq and HIPSD&R-seq

The following sections (5.2.14.1–5.2.14.6) were adapted from the HIPSD-seq publication, to which I contributed as a co-author⁵². UMAP analysis in section 5.2.14.6 was written by Jan Otoničar.

6.2.14.1. Tn5 loading

A high-activity Tn5 transposase³⁰⁰ was used instead of the Tn5 transposase provided in the 10X Multiome kit. Annealing buffer (50 mM NaCl, 40 mM Tris, pH 8) was utilized to resuspend lyophilized adapter oligonucleotides at a concentration of 100 µM. The adapters (Oligo list, table S1) were pre-annealed on a thermocycler by heating at 85°C for 2 min, then gradually cooled to 20°C at a rate of 1°C per minute. For HIPSD-seq⁵², Read1 was annealed with blocked-phos-ME, while Read2 was annealed with blocked-phos-ME. For HIPSD&R-seq⁵², phos-Read1 was annealed with phos-ME, and phos-Read2 with phos-ME. Adapters were mixed with 100% glycerol in a 1:1 ratio and stored at -20°C. Tn5 assembly involved mixing the transposase with the annealed primers in equal volumes, followed by incubation at room temperature for 30 min. The assembled Tn5 was subsequently diluted to a final concentration of 83 µg/ml in dilution buffer (50 mM Tris pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.1% NP-40, 50% glycerol) for HIPSD-seq.

Materials and methods

6.2.14.2. HIPSD-seq and HIPSD&R-seq sample preparation

For HIPSD-seq, cells from LFS041 p.27 (crisis) and LFS041 p.62 (post-crisis) were collected after trypsinization using 0.25% Trypsin/EDTA and cell resuspension in PBS. For HIPSD&R-seq, cells from LFS041 p.22 (early) and LFS087 p.196 (late) were similarly harvested and resuspended in PBS with 1% BSA.

6.2.14.3. Nuclei extraction and nucleosome depletion

For HIPSD-seq, fixation was performed by incubating 1×10^6 cells in 1 ml of 1.5% methanol-free formaldehyde (FA; Thermo Fisher Scientific, Cat. No. 28906) in PBS for 10 min at room temperature while gently shaking. 200 mM glycine was added for fixation neutralization, followed by ice incubation for 5 min. The cells were then centrifuged at 550 g for 5 min at 4°C and washed with ice-cold PBS. To isolate nuclei, the cells were resuspended in 1 ml of ice-cold NIB buffer (10 mM Tris-HCl; pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% Igepal and 1x protease inhibitor cocktail, Cat. No. 5871S, Cell Signalling Technology), followed by incubation on ice with gentle mixing for 20 min. The nuclei were centrifuged at 500 g for 5 min at 4°C and washed once with 1x NEBuffer 2.1 (NEB, Cat. No. B7202). For the nucleosome depletion step, nuclei were resuspended in 1x NEBuffer 2.1 containing 0.3% SDS (Serva, Cat. No. 20767), followed by incubation at 42°C for 15 min with shaking. The SDS was quenched by adding 2% Triton X-100 (Sigma-Aldrich, Cat. No. 93443) and incubating at 42°C with shaking for 15 min. Following centrifugation at 500 g for 5 minutes at 4°C, the nuclei were resuspended in 1x Nuclei Buffer (10X Genomics). Luna-FL™ cell counter (Logos Biosystems) was used to count the nuclei, which were then diluted to a final concentration of 2000–5000 nuclei/μl.

For HIPSD&R-seq, the protocols for fixation with 1.5% FA, nuclei isolation with NIB buffer, and nucleosome depletion using 0.3% SDS were carried out as described above for HIPSD-seq. However, all buffers used during and after nuclei isolation (i.e. to NIB buffer and 1x NEBuffer 2.1) were supplemented with 1 U/μl RNase inhibitor (Takara Bio, Cat. No. 2313A). After nucleosome depletion quenching with 2% Triton X-100, nuclei were centrifuged at 500 g for 5 min at 4°C and resuspended in 1x Nuclei Buffer (10X Genomics), which contains 1 U/μl RNase inhibitor. Finally, Luna-FL™ cell counter was used to count the nuclei, which were ultimately diluted to a concentration of 2000–5000 nuclei/μl.

6.2.14.4. HIPSD-seq library preparation

1.67×10^6 cells from LFS041 p.27 and 1.1×10^6 cells from LFS041 p.62 were used for nuclei isolation and nucleosome depletion as described earlier. The nucleosome-depleted nuclei were processed following the 10X ATAC protocol described in the Chromium Single Cell ATAC Reagent Kits User Guide v1.1 Chemistry (10X Genomics, Cat. No. CG000209), with the modification that the provided transposase was substituted with a highly active in-house Tn5 (83 μg/ml). For transposition, 10,000 nuclei were loaded onto the Chromium Next GEM Chip

H (PN-1000161). Single-cell DNA libraries were prepared using the standard reagents of the Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit v1.1 (PN-1000176), with unique indexing applied via the Single Index Kit N Set A (PN-1000212). Qubit 3.0 Fluorometer (Invitrogen, Cat. No. Q33216) and the 4200 TapeStation system (Agilent Technologies) were used for quality control and to perform molarity calculations of the final libraries. Sequencing of the libraries was carried out using the NovaSeq 6000 platform with paired-end 100 SP reads: 51 cycles for read 1, 51 cycles for read 2, 8 cycles for i7, and 16 cycles for index 2, with 1% PhiX spike-in.

6.2.14.5. HIPSD&R-seq library preparation

A total of 4.3×10^5 cells from LFS041 p.22, while 5×10^5 cells each from LFS041 p.67 and LFS087 p.196 were pooled (1:1 ratio)⁵² for the mixed patient experiment for HIPSD&R-seq were used for nuclei isolation and nucleosome depletion as described above. 10X Multiome protocol, according to the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Kits User Guide (10X Genomics, Cat. No. CG000338) was used to process the nucleosome-depleted nuclei.

For transposition, 10,000 nuclei were loaded onto the Chromium Next GEM Chip J (PN-1000230), with the 10X Genomics transposase replaced by a highly active in-house Tn5. The standard reagents from the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle (PN-1000285) were used to prepare both scDNA and scRNA libraries. The scDNA libraries were indexed with the Single Index Kit N Set A (PN-1000212), whereas the scRNA libraries were indexed with the Dual Index Kit TT Set A (PN-1000215). Qubit 3.0 Fluorometer (Invitrogen, Cat. No. Q33216) and the 4200 TapeStation system (Agilent Technologies) were used for quality control and to perform molarity calculations of the final libraries. Sequencing of scDNA libraries was performed on the NovaSeq 6000 platform (200 cycles, S4) with 101 cycles for read 1, 101 cycles for read 2, 8 cycles for i7, and 24 cycles for i5. The NovaSeq 6000 platform (paired-end 100 SP) was used to sequence scRNA libraries, with 28 cycles for read 1, 90 cycles for read 2, and 10 cycles each for i7 and i5 indices. All libraries were sequenced using 1% PhiX.

6.2.14.6. HIPSD-seq and HIPSD&R-seq computational analysis

The initial preprocessing of HIPSD-seq (crisis and late passage) and HIPSD&R-seq (early passage and the mixed experiment) data was performed with Cell Ranger ATAC pipeline (version 2.1.0), with a reference provided by 10X Genomics (Cell Ranger Arc, GRCh38, version 2020-A). In the case of HIPSD&R-seq, only the DNA component was analysed. Because the default cell definition from Cell Ranger is not appropriate for samples with depleted nucleosomes, the cells were defined based on the barcode rank plot ($\log(\text{fragments})$ vs. $\log(\text{rank})$). Only cells above the highest gradient were used for further processing.

Materials and methods

The initial BAM output from Cell Ranger was then split into single-cell BAM files with a command 'bamsplice' from Cell Ranger DNA pipeline (version 1.1.0). Next, single-cell BAM files were filtered to retain only high-quality reads with SAMtools²⁹¹. To extract the number of counts per 1MB bin size, hmmcopy_utils was used³⁰¹. Only cells with more than 90% of non-empty bins and 60,000 counts were kept for copy number calling with HMMcopy²³⁸. When running HMMcopy, the e value was set to 0.9999. Because we previously observed a tendency of HMMcopy to overcall homozygous deletions in our data, we encoded both, homozygous and heterozygous deletions as a loss with a copy number of 1. Finally, CNAs were plotted with ComplexHeatmap²³³, only using genomic bins that are annotated as ideal by HMMcopy.

For the mixed-patient (LFS041 p.67: LFS087 p.196 = 1:1) HIPSD-seq experiment, the data was processed as described above. However, the next approach was used to select only the cells from the patient LFS087 p.196. Single-cell CNA profiles of all the cells were correlated (Pearson's correlation) to bulk CNA profiles for both patients. In order for a cell to be classified as LFS087 p.196, the cell needed to have higher correlation to bulk CNA profile of the patient LFS087 than to CNA profile of the patient LFS041.

UMAP (Uniform Manifold Approximation and Projection) was generated with the use of scanpy²⁶², PCA was performed on the combined CNA matrix of all LFS041 three samples (early, crisis and post-crisis) and first 15 PCs were used to construct a KNN (K-Nearest Neighbours) graph. Finally, UMAP was used to visualize the three clusters in two dimensions, with points being coloured by sample.

6.2.15. Strand-seq

6.2.15.1. Strand-seq library preparation and sequencing

Strand-seq libraries were prepared following the protocol detailed previously⁵⁴ with minor modifications. The key steps of the procedure are summarized below. In brief, BrdU incorporation was carried out using growing cells from three distinct passages from LFS041 (p.22 (early), p.63 (post-crisis) and p.343 (late)). Cells were labeled with 40 μ M BrdU (Sigma, B5002) for a single round of cell division. This step is crucial as cells with incomplete BrdU incorporation or those that have undergone multiple DNA synthesis phases with BrdU present will yield unreliable strand-specific sequencing data. Cells were frozen and stored at -80°C until further processing.

Upon thawing in supplemented MEM medium, cells were centrifuged and resuspended in Nuclei Staining Buffer A to a final concentration of 1×10^6 cells/ml. The buffer composition was as follows: 100 mM Tris-HCl (pH 7.4), 154 mM NaCl, 1 mM CaCl₂, 0.5 mM MgCl₂, 0.2% BSA, 0.1% NP40 (Sigma-Aldrich, Cat. No. 74385), 10 μ g/ml Hoechst 33258 (Enzo Life Sciences, Cat. No. ENZ-52402) in ultra-pure water. The cell suspension was filtered through a cell

strainer and incubated on ice for approximately 30 min. Single cells were then sorted into individual wells of a 96-well plate using fluorescence-activated cell sorting (FACS), each containing 5 µl of TheraPEAK ProFreeze™ Freezing Medium (Lonza, Cat. No. BEBP12-769E) and the plate was stored at -80°C until further use. Following thawing, DNA MNase fragmentation was carried out by incubating the samples with 0.5 U (Unit) Micrococcal Nuclease (NEB, Cat. No. M0247S) in MNase buffer supplemented with 1.5 mM DTT (Sigma-Aldrich, Cat. No. 43816) and 5% PEG 6000 (Calbiochem, Cat. No. 528877) for 8 min at room temperature, in a total reaction volume of 15 µl/well. The enzymatic reaction was terminated by adding 10 mM EDTA. DNA purification was performed using AMPure XP beads (Beckman Coulter, Cat. No. A63881) at a 1.0X ratio, with subsequent elution in 10 µl EB Buffer, utilizing a Biomek FXp liquid handling robotic system to facilitate large-scale library preparation.

End-repair was carried out by incubating the samples for 30 min at room temperature with a mixture containing 0.3 U/µl T4 DNA polymerase (NEB, Cat. No. M0203S), 0.1 U/µl Klenow DNA polymerase (NEB, Cat. No. M0210S), and 1 U/µl T4 polynucleotide kinase (NEB, Cat. No. M0201S), 1x T4 ligase buffer (NEB, Cat. No. B0202S) and 2 mM dNTP mix (NEB, cat. no. N0447S), whereas 2.5 µl of the mixture was added to each well. Following this step, the DNA again underwent clean up using AMPure XP beads at a 1.8X ratio, employing the Biomek FXP liquid handling robotic system to maintain consistency and efficiency. Next, A-tailing was performed by adding 1.5 µl of A-tailing Master Mix to each well, which contained 1.67 U/µl Klenow Fragment (3'→5' exo-; NEB, Cat. No. M0212S) and 1.67 mM dATP (NEB, Cat. No. N0440S). The samples were incubated at 37°C for 30 min, then purified with AMPure XP beads at a 1.8X ratio. Subsequently, adapter ligation was performed by incubating DNA fragments with forked Illumina adaptors in a master mix prepared using 0.0893 µM PE adaptors, 2.67x T4 Quick Ligation Reaction Buffer, and 0.144 U/µl Quick Ligase (NEB, Cat. No. M2200L; with the included buffer), diluted in Ultrapure H₂O to a final volume of 7.5 µl per well. After adding the master mix to the DNA fragments, the reaction volume reached 20 µl, resulting in a final adaptor concentration of 33.5 nM per cell. The ligated DNA was purified once more using AMPure XP beads at a 1.6X ratio and eluted in a volume of 9.5 µl.

Subsequently, the DNA was incubated with 10 µg/ml Hoechst 33258 for 15 min, followed by UV irradiation. The irradiation was carried out using a crosslinker equipped with five 365-nm longwave UV bulbs for 15 min, delivering a total dose of 2.7×10^3 J/m². The nicked DNA was amplified using a combination of 96-well-custom multiplexing Primer PE 2.0 which includes a combination of oligonucleotides each containing a 6-bp (hexamer) multiplexing barcodes (Sigma-Aldrich), and Primer PE 1.0 (Illumina). First, 1 µl of Primer PE 2.0 was added per well, followed by 12.5 µl of Phusion HF PCR Master Mix (NEB, Cat. No. M0531L) and 1 µl of Primer PE 1.0, bringing the total reaction volume to 25 µL. Thermal cycling conditions were as follows: initial denaturation at 98°C for 30 seconds, followed by 18 cycles of denaturation at 98°C for

Materials and methods

10 sec, annealing at 65°C for 30 sec, and extension at 72°C for 30 sec. A final extension step was carried out at 72°C for 5 min. Post-amplification, the DNA from all wells was pooled and subjected to purification and size selection using AMPure XP beads at a 0.8X ratio. This step served to remove free primers and adapter dimers. The resulting libraries were sequenced on a NextSeq500 platform using a MID-mode, 75bp paired-end protocol. Following sequencing, the data was demultiplexed, aligned to the hg38 reference assembly using BWA mem (0.7.17-r1188)²⁸⁹, and duplicate reads were mapped using sambamba (v1.0)²⁷³.

6.2.15.2. Strand-seq data processing

Strand-seq data was processed using the MosaiCatcher pipeline (v2.0.1)³⁰². Mapping quality filtered (MAPQ > 10) and non-duplicated reads were used to generate read counts per 200kb bin for Watson and Crick strands separately in each cell. Read counts per bin were then visualized per chromosome and cell. Prior to downstream analysis, cells were manually filtered based on their visualized read count distributions to exclude noisy cells (those exhibiting spiky peaks, high background or uneven coverage), cells with incomplete BrdU incorporation, cells with fewer than 200,000 reads, and those displaying all-Watson or all-Crick chromosomes or extremely skewed Watson/Crick ratios (>80% Watson or >80% Crick). For the remaining cells, the read counts per strand and chromosome were segmented, haplotype-resolved using StrandPhaseR and fed into MosaiClassifier to create initial SV calls per cell.

Ideograms generated using MosaiCatcher were used to profile SVs, including deletions, duplications, inverted duplications, inversions, and complex rearrangements, in single cells across different passages using Microsoft Excel. Each SV was mapped to its genomic position in Mb resolution for individual chromosomal arms after detailed examination of each cell. This comprehensive approach facilitated variant calling and enabled us to create a matrix of all possible SVs for each cell.

6.2.15.3. Clustering of single cells based on SVs

This section was primarily written by Petr Smirnov and adapted by me with minor additions. Next, two approaches to cluster single cells based on SV annotations per-cell detected from StrandSeq were taken (as described above). The first approach was event-based. For this approach, unique breakpoints detected across all cells were used to segment the genome. For each cell, the presence or absence and the type of SV at each segment was recorded. The second approach was bin-based. Here the genome was binned into approximately 2 MB bins, with equal bin-sizes per chromosome derived from the tile function as implemented by the GenomicRanges R package²⁴³. For each cell and segment derived from binning, an SV was recorded if it was detected by Strand-seq to overlap that segment within that cell. In the case of two or more SVs overlapping the same segment, the SV overlapping the greater number of

bases within the segment was assigned. In both cases, a virtual diploid cell which is unmodified at each genomic segment was added for subsequent rooting of phylogenetic trees.

For both genome segmentation approaches, distance between cells was then calculated using a modified Hamming distance, where for each pair of cells, each segment contributed a distance of 1 if the event recorded for that segment did not match, except for a comparison between unmodified and complex segments, which contributed a distance of 2. This was chosen to account for the fact that the “complex” SV annotation as described above can arise from the overlap of multiple simple SVs in the same genomic location. The cells were then clustered using the neighbour joining method as implemented by the ape R package²²⁶. For visualization purposes, the ape package was used to plot the resulting clustering as phylogenetic trees, rooted at the virtual diploid cell added to the data. Finally, main structural variance differences and shared events between main clusters were annotated manually based on the matrices created earlier using Excel.

6.2.15.4. Fold-back inversions calling (BFB patterns) in Strand-seq data

I analysed the Strand-seq data to identify fold-back inversions in single cells, which are hallmark features of BFB events⁵⁵. Fold-back inversions were quantified by searching for characteristic patterns on all chromosomal arms, defined as small amplifications involving either the H1 (haplotype 1), H2 (haplotype 2) strand or both, followed by a deletion. The classification of fold-back inversions and their connection to BFB cycles was guided by previously defined structural rearrangement footprints⁵⁵. GraphPad Prism 8 software was used to create graphs, perform statistical tests and calculate p-values. Fisher's exact test was performed for pairwise comparisons between each of the three passages. To account for the number of passages ($n = 3$), results were normalized accordingly, and p-values were adjusted as comparisons were made across the three passages.

6.2.16. Primary Template-directed Amplification (PTA)

PTA sequencing was performed on the same passage (p.33) for both *hTERT*⁺ cells and pMX-GFP control cells. Cells were washed, trypsinized and centrifuged as previously described. The resulting cell pellets were resuspended in 200-300 μ l of PBS supplemented with 5% FCS, depending on the size of the pellet. Prior to cell sorting, cells were stained with 0.1 μ g/ml DAPI (Roche, Cat. No. 10236276001) to assess viability.

6.2.16.1. FACS sorting

Cell sorting was performed using a BD FACSymphony S6 Cell Sorter (BD) DKFZ Flow Cytometry Core Facility, with BD FACSDiva 9.8 software for data acquisition and analysis. FSC-A (side scatter area) vs. SSC-A (forward scatter area) gating was applied to identify cells and exclude debris, while SSC-A vs. SSC-W (side scatter width) gating was used to

Materials and methods

discriminate single cells from doublets and clumps. Finally, 431/28 vs. SSC-A gating was applied to distinguish DAPI-positive cells, allowing the exclusion of dead or permeabilized cells. Cells were sorted into twin.tec PCR plate 384-well Lobind plates (Eppendorf, Cat. No. EP0030129547), prefilled with 0.5 µl of cell lysis buffer (Bioskryb) at the DKFZ Single-Cell Open Lab. One half of the plate was sorted with *hTERT*⁺ cells, while pMX-GFP control cells were sorted into the other half. As controls, one well in each half contained 10 cells and another well was left empty. Following sorting, cells were centrifuged at 150g for 1 minute and placed directly on dry ice and stored at -80°C until further processing.

6.2.16.2. Whole genome amplification using PTA protocol

ResolveDNA Whole Genome Amplification Kit v2.0 (Bioskryb) was used for single-cell genome amplification. The protocol was based on the original Bioskryb protocol and adapted by the DKFZ Single-Cell Open Lab.

Prior to amplification, the plate was thawed on ice for 5 min and centrifuged shortly before further use. The 12X SS2 reagent, control gDNA, DNA/Nuclease-free water and Cell Buffer were thawed at room temperature, while the reagents SM3, SDXT and SB5 were placed on ice to thaw for approximately 30 min. All reagents were vortexed briefly and spun down before use. A MS lysis mix was prepared on ice according to Table 7.

Table 7. Composition of MS lysis mix. *20% overage was included.

Component	Volume per well (µl)	Volume per 384 reactions (µl)*
SDXT Reagent	0.6	276.5
12X SS2 Reagent	0.025	11.5
SM3	0.375	172.8
Total volume	1	460.8

Next, 1 µl of the MS lysis mix was added to each well using the Mantis liquid dispenser (Formulatrix). The plate was then sealed, briefly centrifuged for 10 sec, and mixed on a Bravo Teleshaker (Agilent Technologies) at 1400 rpm for 20 min. The amplification reaction mix was then prepared on ice as shown in Table 8.

Table 8. Composition of the amplification reaction mix. *20% overage was included.

Component	Volume per well (µl)	Volume per 384 reactions (µl)*
SB5 Reagent	1.85	852.5
SEZC Reagent	0.15	69.1
Total volume	2	921.6

Afterwards, the Mantis liquid dispenser was used to add 2 µl of the amplification reaction mix was added to each well. The plate was then also sealed, spun down shortly for 10 sec and mixed on the Bravo Teleshaker for 1 min at 1000 rpm. The plate was spun down again for 10

sec before starting the DNA amplification protocol on the thermal cycler. The protocol was as follows (total volume: 4 µl): lid temperature at 70°C, followed by 30°C for 150 min, 65°C for 3 min, and then cooled down to 4°C. To assess the amplification, Qubit was used to measure the concentration of DNA in 8 random wells for each condition (*hTERT*⁺ and pMX-GFP control cells). Additionally, an Agilent TapeStation electropherogram was used to analyse the fragment size. The plate was then stored at -20°C overnight.

6.2.16.3. PTA library preparation and sequencing

The PTA library preparation protocol for a 384-well plate was provided by the DKFZ Single-Cell Open Lab. The plate containing genomic DNA (gDNA) was thawed on ice for approximately 30 minutes before processing. For tagmentation, a Tn5 master mix was prepared by adding 0.4 µl Tn5³⁰⁰ (1:20 dilution) to 0.4 µl DMF (Thermo Fisher Scientific, Cat. No. 20673) and 0.4 µl of 4x tagmentation buffer (40 mM Tris-HCl pH 7.5, MgCl₂ in H₂O) to each well of a 384-well plate using the Mantis liquid dispenser. The Mosquito LV (SPTLabtech) was then used to transfer 0.4 µl gDNA from each well into the plate containing the Tn5 master mix. The plate was subsequently sealed, briefly centrifuged, and incubated in a 384-well with a lid temperature of 70°C, followed by incubation at 55°C for 10 min before cooling to 4°C. To inactivate Tn5, 0.4 µl of 0.1% SDS was added to each well using the Mantis liquid dispenser, and the plate was kept on ice for 5 minutes. Barcoding of individual cells was performed by adding 0.4 µl of 3 µM Nextera DNA Indexes (i5: 501–516, i7: 701–724) into each well using the Mosquito. Next, 2.7 µl of 2x KAPA HiFi HS ReadyMix and 0.3 µl of DMSO were added to each well using the Mantis liquid dispenser. The plate was subsequently sealed, vortexed, spun down, and subjected to PCR amplification in a 384-well thermocycler using the following program: 72°C for 3 minutes, 98°C for 30 seconds, followed by 12 cycles of 98°C for 20 seconds, 63°C for 15 seconds, and 72°C for 30 seconds. A final extension was performed at 72°C for 3 minutes before holding at 4°C.

For library clean-up, the libraries (*hTERT*⁺ p.33 or pMX-GFP and p.33) were pooled separately, resulting in two distinct libraries. Each library was purified using AMPure XP beads at a 0.9X ratio and incubated for 5 min at room temperature. The samples were then placed on a magnetic separator until the solution cleared, after which the supernatant was discarded. The bead pellets were washed twice with 80% ethanol, followed by ethanol removal while the tubes remained on the magnetic separator. Residual ethanol was collected by centrifugation and removed, and the beads were left to air dry for approximately 3–5 minutes. DNA was then eluted with 70 µl of elution buffer (Qiagen, Cat. No. 19086) for 3 minutes, placed on the magnetic separator, and the supernatant containing the final libraries was transferred to LoBind Eppendorf tubes for storage at -20°C until sequencing. Library concentrations were measured using Qubit, while fragment sizes were assessed using the D1000 Tape on the TapeStation

Materials and methods

4200 electropherogram (Agilent Technologies). Sequencing was performed on the NovaSeq 6000 platform with paired-end 100 SP reads at the DKFZ NGS Core Facility.

6.2.16.4. PTA bioinformatics analysis

This section was primarily written by Jan Otoničar. The FASTQ files were first aligned to the human reference genome (GRCh37) for each cell separately with BWA-MEM2 (version 2.2.1)³⁰³. Next, single-cell BAM files were filtered to remove poor-quality reads with SAMtools²⁹¹. The clean BAM files were next used as an input for scAbsolute³⁰⁴, using 500kb bin size genomic windows and limiting possible ploidy solutions between 1.2 and 2.9. Single-cell CNA profiles were manually inspected to retain only high-quality cells. In order to calculate the fraction of genome altered per cell, the number of bins not equal to 2 was divided by the total number of bins used by scAbsolute. The null hypothesis that the fraction of genome altered per cell is sampled from the same distribution was tested using the Wilcoxon rank-sum test in the Python package SciPy²⁶³. Data visualization, including the boxplot, was performed using Python libraries, Matplotlib²⁶⁰ and seaborn²⁶⁴.

6.2.17. Analysis of hypertranscription

I re-analysed data from a study on hypertranscription, which encompasses RNA output levels from 7494 patients from The Cancer Genome Atlas (TCGA) across 31 cancer types¹⁶⁴. The authors calculated the hypertranscription fold change using the RNAmP method, which compares the allelic ratio of somatic mutations in RNA to their corresponding ratio in DNA, leveraging the increased representation of transcribed alleles in RNA compared to DNA. The patient IDs of the same 7494 patients were used to classify them into two groups based on *TP53* and *MDM2* status from the publicly available TCGA dataset through the Genomic Data Commons (GDC) portal. The first group included tumours with WT *TP53* and *MDM2* (WT *TP53/MDM2*), while the second group consisted of tumours with *TP53* loss, *MDM2* gain, or both (*TP53/MDM2*⁺/*TP53MDM2*⁺), which are referred to *TP53/MDM2*⁺ in the main text. Both heterozygous and homozygous CNAs were considered due to the low frequency of homozygous *TP53* copy number losses and *MDM2* copy number gains. After filtering, the dataset contained of 6,305 tumour samples, split into 4032 *TP53/MDM2*⁺ and 2273 WT *TP53/MDM2* tumours.

The fold increase in RNA output, namely the hypertranscription fold change (hytx_fch) was used to compare the hypertranscription status between the two groups. The comparison was performed based on the total number of patients, regardless of the tumour type, as well as for each tumour type separately.

Analysis of hytx_fch based on the WT status of *MYC* was performed after stratifying the cohort into *TP53/MDM2*⁺/WT *MYC* and WT *TP53/MDM2/MYC* groups for comparative analysis. For

comparisons within each cancer type, a minimum sample size of 50 patients per group was required to ensure statistical robustness. Tumour types that did not meet this criterion were excluded from the analysis. GraphPad Prism 8 software was used to create graphs, perform statistical tests and calculate p-values. Statistical significance was assessed using a non-parametric t-test (Mann-Whitney test) to compare two independent groups, as the data were not normally distributed. P-values less than 0.05 were considered statistically significant (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

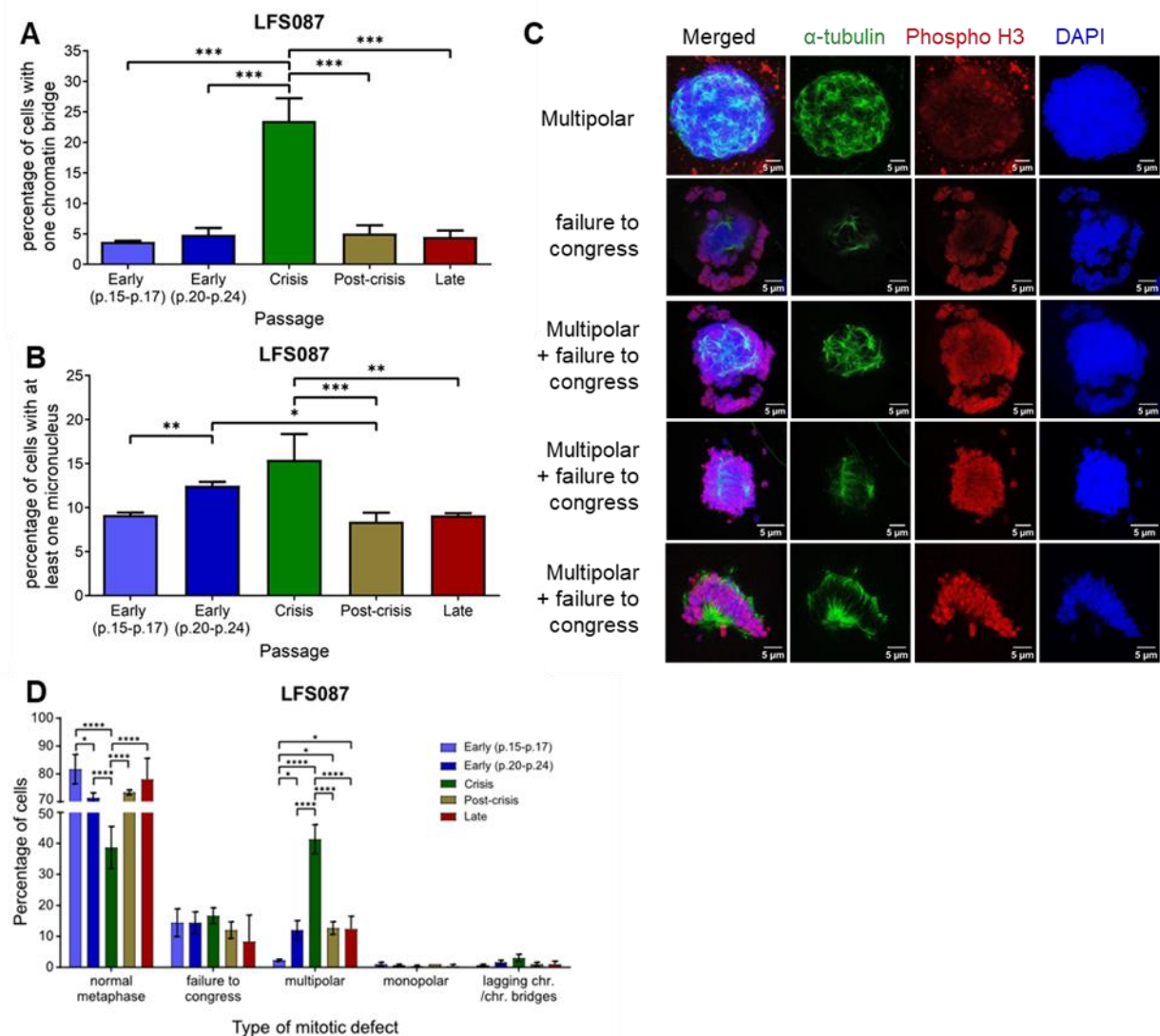
6.2.18. Statistics and reproducibility

Statistical analyses were performed using the tests described in the respective sections and specified in the figure legends or the main text, as appropriate. Statistical significance was defined as a p-value less than 0.05, and multiple testing corrections were applied when needed. Sample sizes were not predetermined using statistical methods, and all collected data were included in the analyses. Normal distribution was assumed for most data. An exception was made for *TERT*-FISH, ALT-FISH and TCGA data, which did not follow a normal distribution. Detailed information on biological replicates and experimental repetitions is provided in the respective figure legends. Experimental replicates varied by assay as indicated. Graphs, statistical analyses, and p-value calculations were conducted using GraphPad Prism 8 software or R studio, unless otherwise stated. All values are reported as the mean $\pm$ standard error of the mean (SEM) of biological replicates unless stated otherwise. All figures and illustrations presented in this thesis were formatted and edited by myself using Affinity Designer, unless indicated otherwise.

6.2.19. Thesis writing

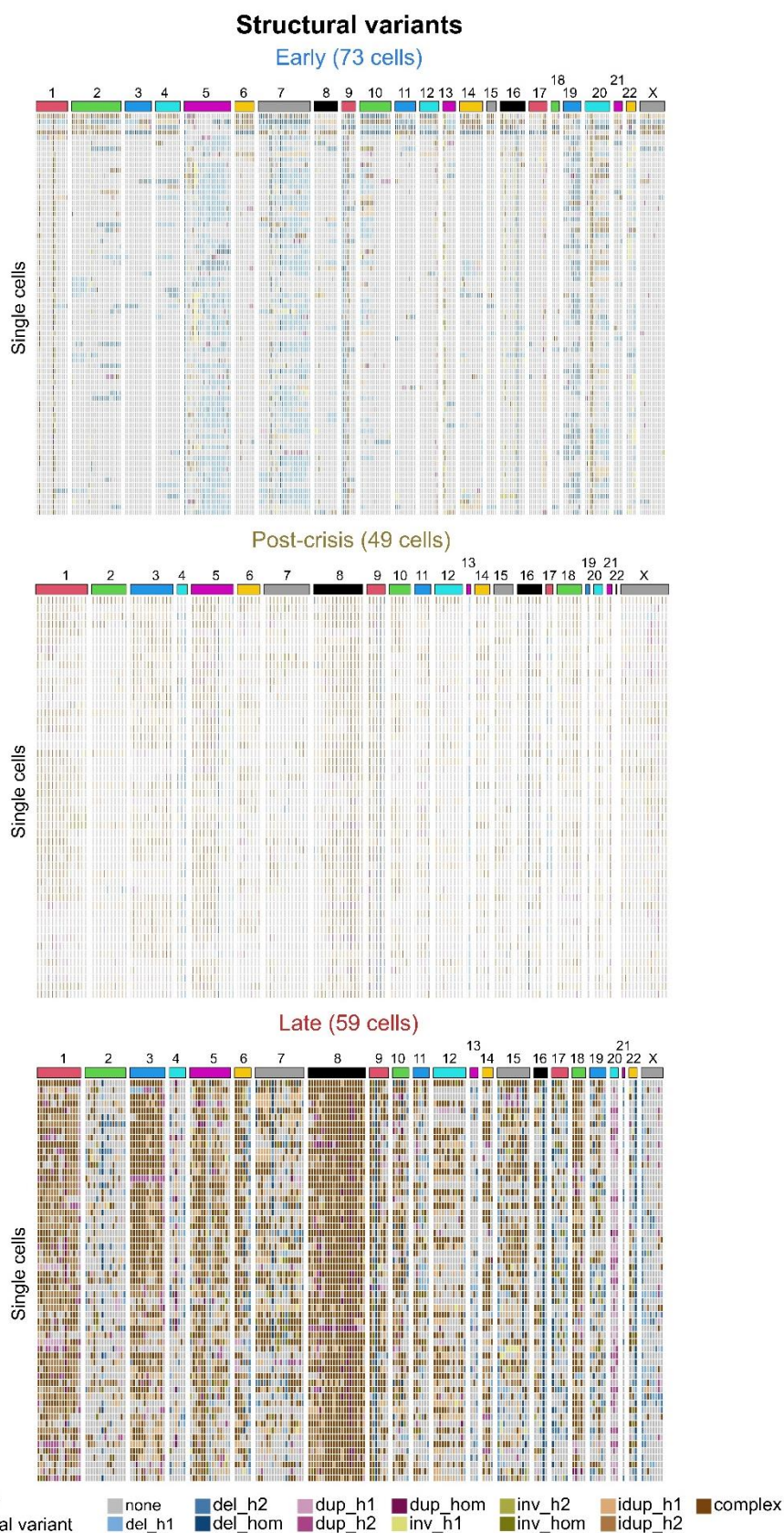
The text in this thesis is my original work. I used ChatGPT and Perplexity for grammar correction and readability improvements, but all content, ideas, and arguments are entirely my own. The abstract was translated into German with the help of DeepL.

Supplementary figures



Supplementary Figure 1. Phenotypic characteristics of LFS skin-derived fibroblasts.

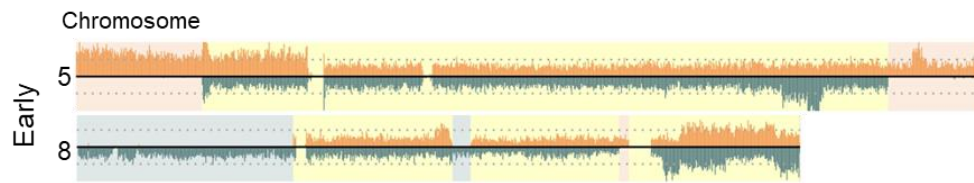
A. Quantification of chromatin bridges in patient LFS087. Three independent biological replicates were performed for each condition, with 750 cells quantified per replicate. **B.** Quantification of micronuclei in LFS087. Three independent biological replicates were performed for each condition, with 750 cells quantified per replicate. **C.** Representative immunofluorescence images showing examples of mitotic defects. Scale bar: 5 μ m. **D.** Quantification of mitotic defects in LFS087. Three independent biological replicates were performed for each condition, with 100 cells quantified per replicate. Lagging chr./chr. bridges: lagging chromosomes/ chromatin bridges. In all experiments, the same passages were used for each time point: early (p.15–p.17, p.20–p.24), crisis (p.30–p.33), post-crisis (p.47–p.56), and late (p.192–p.201). Data in panels **A** and **B** are presented as mean $\pm$ SEM. Statistical significance was assessed using a one-way ANOVA followed by Tukey's multiple comparisons test. In **D**, data are presented as mean values. Statistical significance was assessed using repeated-measures two-way ANOVA followed by uncorrected Fisher's LSD for multiple comparisons. P-values below 0.05 were considered statistically significant ($n = 3$ per group, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I).



Supplementary Figure 2. SV profiles across passages in LFS041 using Strand-seq. Heatmaps showing all SVs on each chromosome, detected in cells from each passage: early (p.22), post-crisis (p.63) and late (p.343). Each row shows one cell, and each column represents one chromosome. Del: deletion, dup: duplication, inv: inversion, idup: inverted duplication, h1: homolog 1, h2: homolog 2, hom: homologous. Figure adapted from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Petr Smirnov performed the bioinformatics analysis.

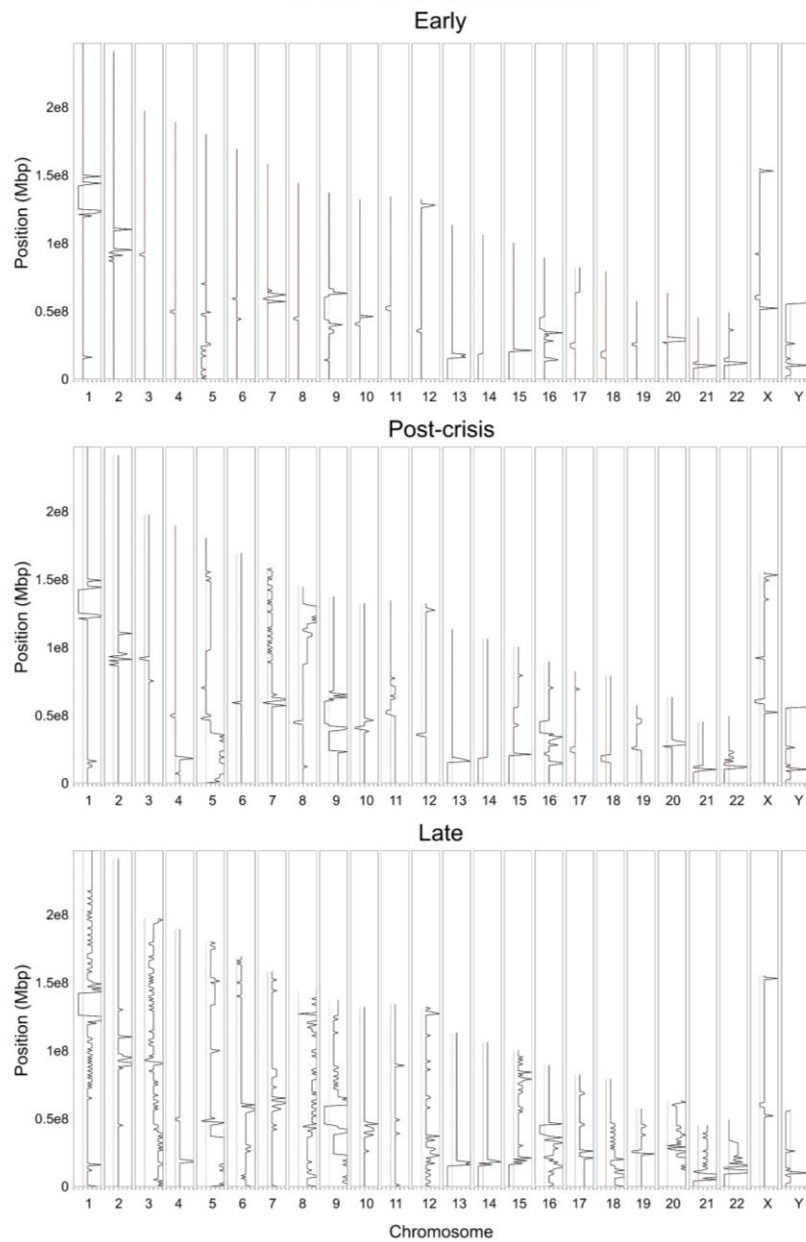
Supplementary figures

Strand-seq data with complex rearrangements similar to post-crisis and late passages

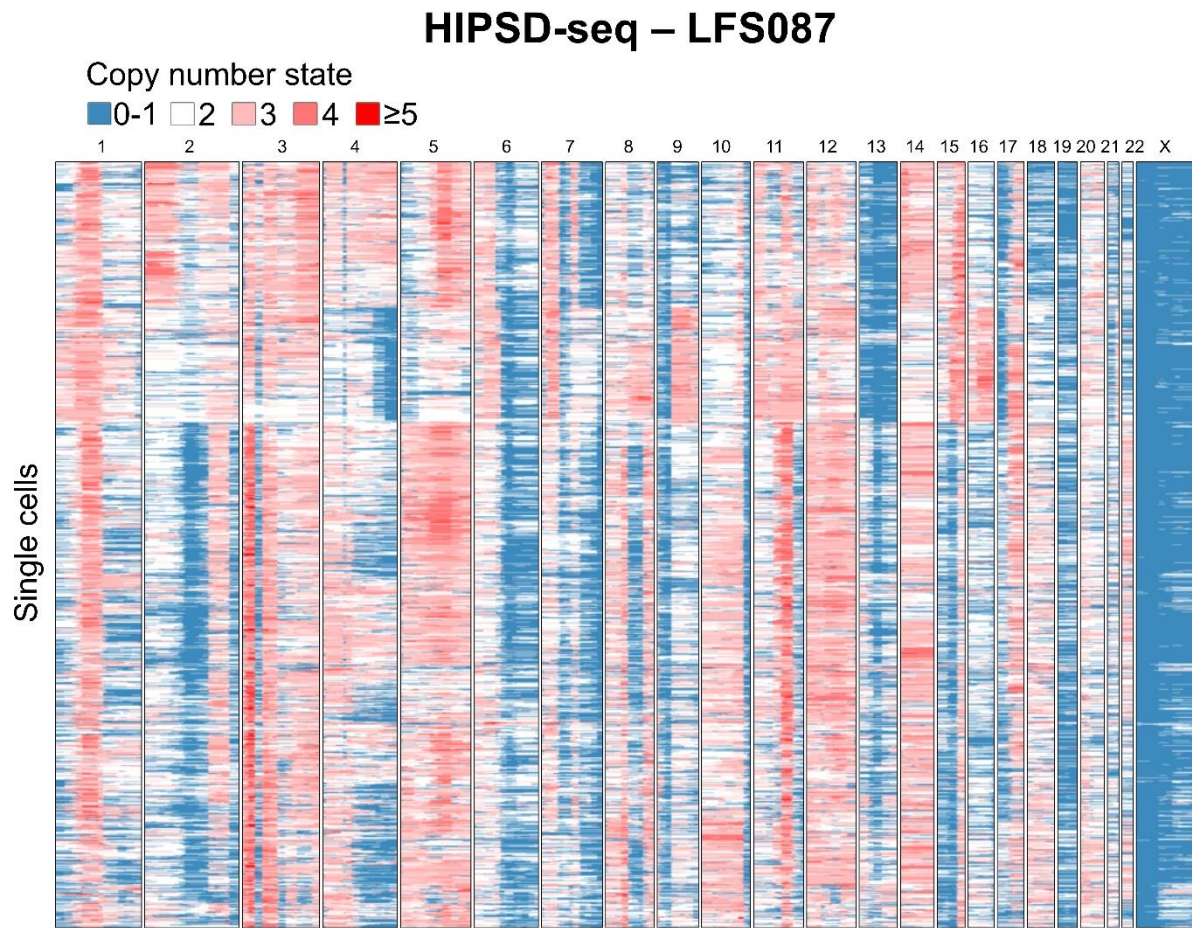


Supplementary Figure 3. Representative Strand-seq examples of chromosomes 5 and 8 in early passage cells (LFS041 p.22). These chromosomes exhibit complex rearrangements similar to those detected as clonal in post-crisis and late passages (Figure 15B-C). Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Petr Smirnov performed the bioinformatics analysis.

Strand-seq – pseudobulk

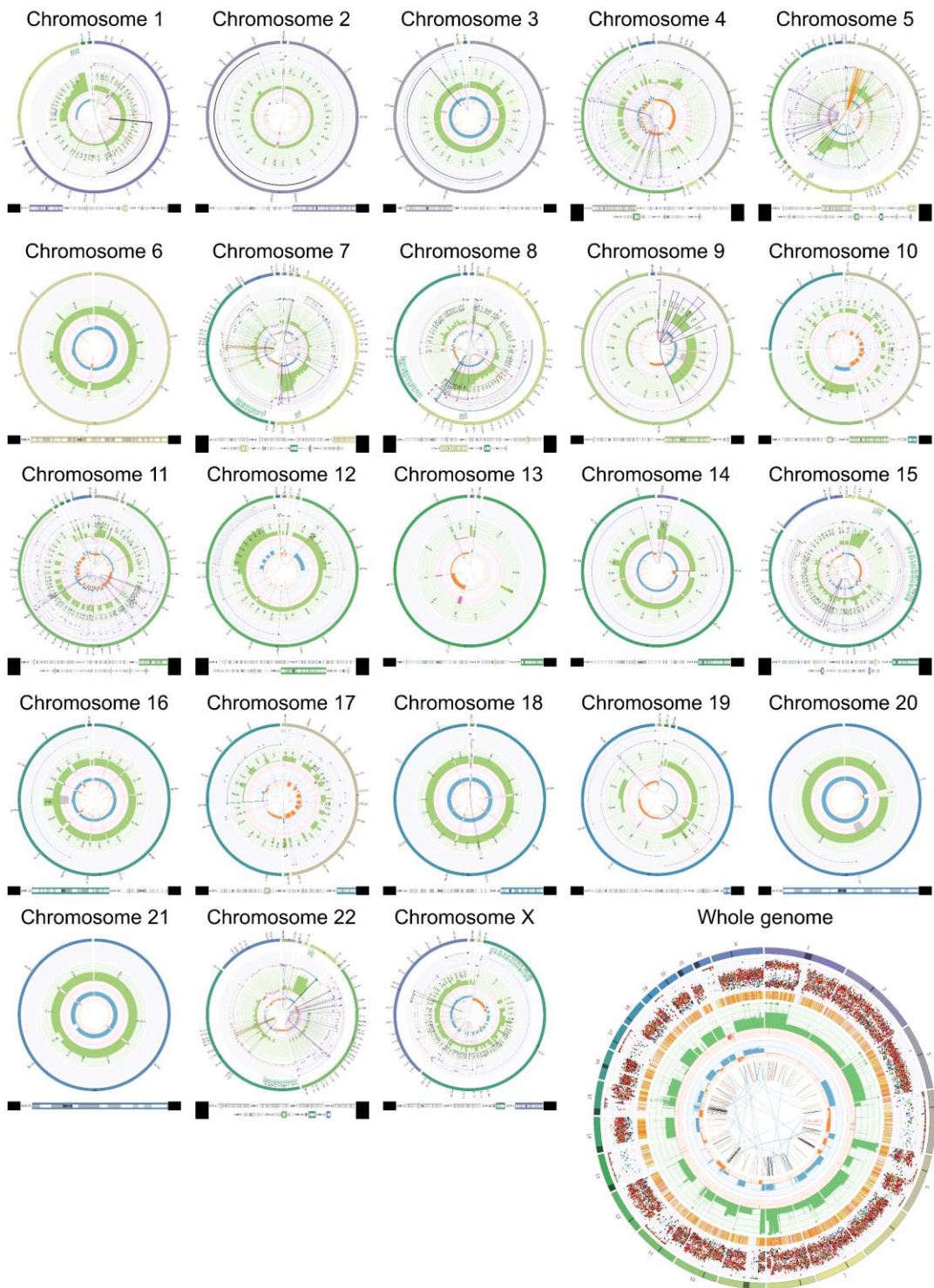


Supplementary Figure 4. Strand-seq pseudobulk analysis of LFS041. The results show the ploidy status of three different passages: early (p.22), post-crisis (p.63), and late (p.343). The x-axis represents the chromosomes, while the y-axis shows the position in Mbp resolution. The red dashed line indicates the diploid state. Petr Smirnov performed the bioinformatics analysis.

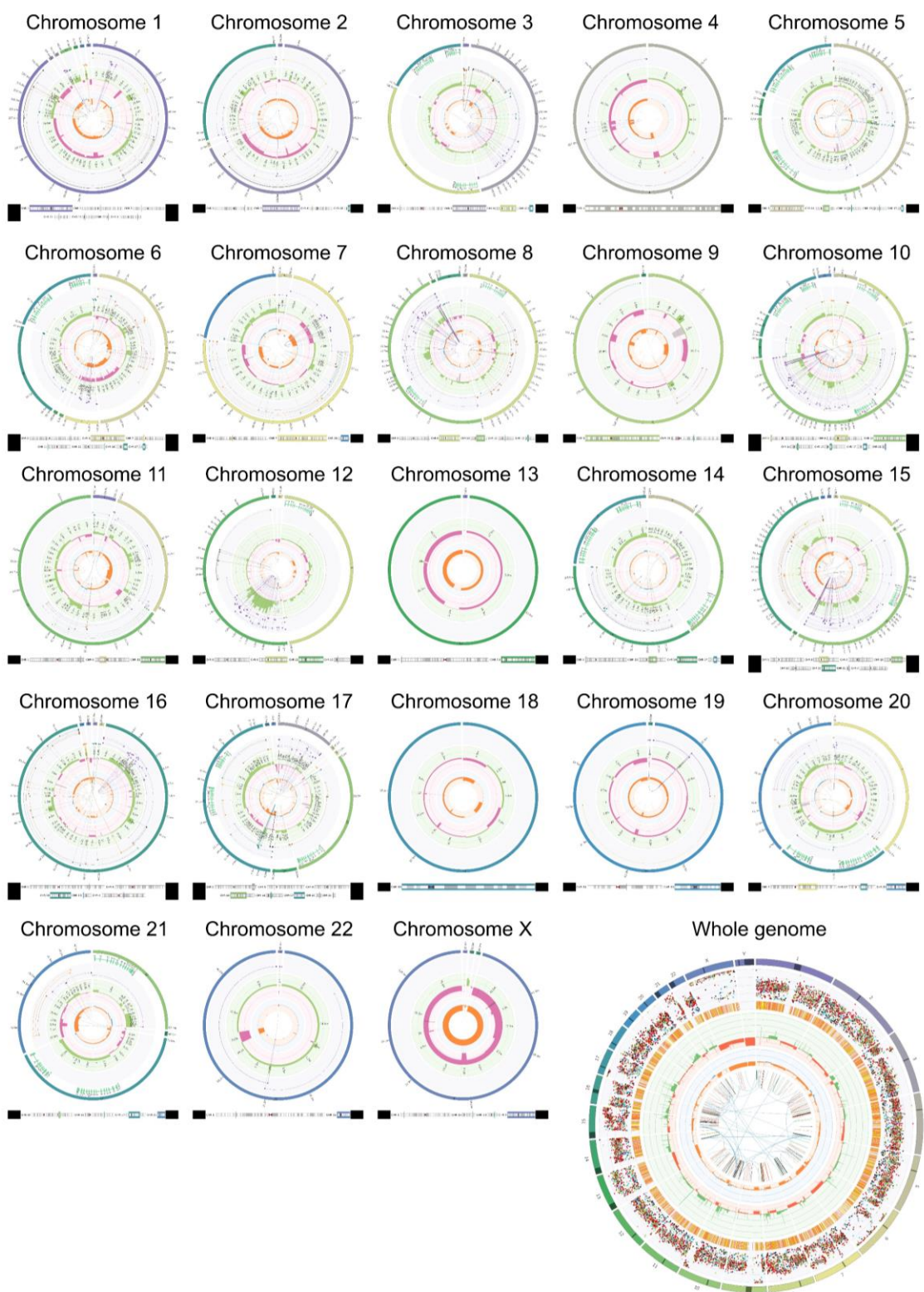


Supplementary Figure 5. Heatmap representation of copy number profiles in LFS087 analysed using HIPSD-seq. Late-passage cells (p.196, n = 1869) are displayed. Each row represents one cell, and each column corresponds to a chromosome with a bin size of 1000 kb resolution. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Pooja Sant performed the experimental part, and Jan Otoničar performed the bioinformatics analysis.

Bulk WGS – LFS041 p.63 (post-crisis) – Circos plots

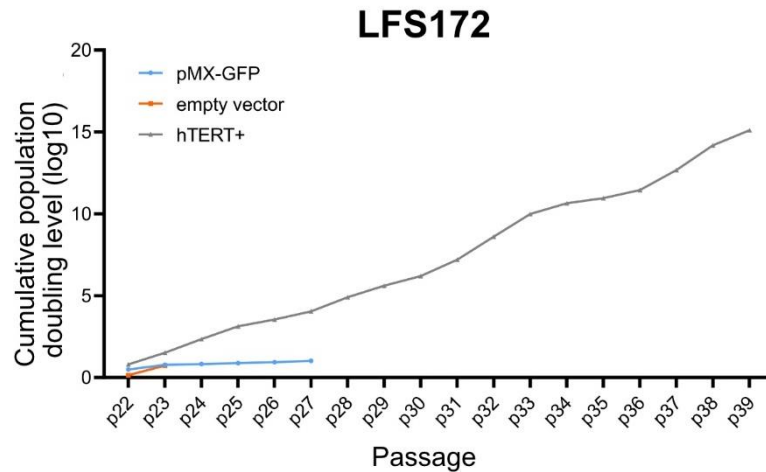


Supplementary Figure 6. Circos plots depicting somatic alterations identified from WGS data of LFS041 p.63 (post-crisis) using OncoAnalyser. Each plot illustrates the distribution of somatic mutations, copy number variations and SVs across chromosome and the whole genome. In the whole-genome Circos plot, the outer track shows the chromosomes, followed by somatic variants (SNP/allele frequency), somatic variants (short indel locations), copy number changes and the inner circle exhibits minor allele copy numbers, while the inner connections indicate relationships between chromosomal regions affected by SVs. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Petr Smirnov performed the bioinformatics analysis.

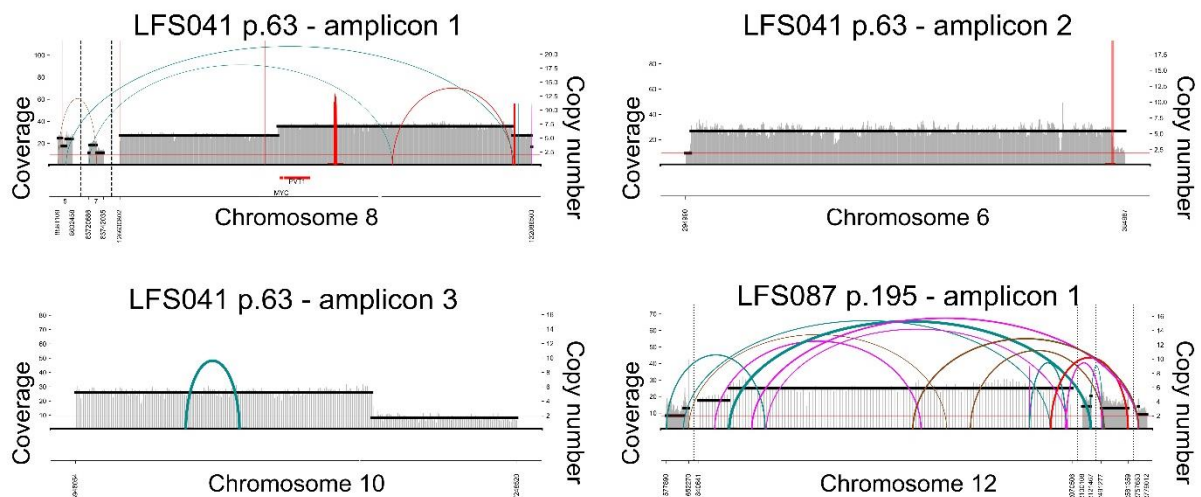
Bulk WGS – LFS087 p.195 (late) – Circos plots

Supplementary Figure 7. Circos plots depicting somatic alterations identified from WGS data of LFS087 p.195 (late) using OncoAnalyser. Each plot illustrates the distribution of somatic mutations, copy number variations and SVs across chromosome and the whole genome. In the whole-genome Circos plot, the outer track shows the chromosomes, followed by somatic variants (SNP/allele frequency), somatic variants (short indel locations), copy number changes and the inner circle exhibits minor allele copy numbers, while the inner connections indicate relationships between chromosomal regions affected by SVs. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Petr Smirnov performed the bioinformatics analysis.

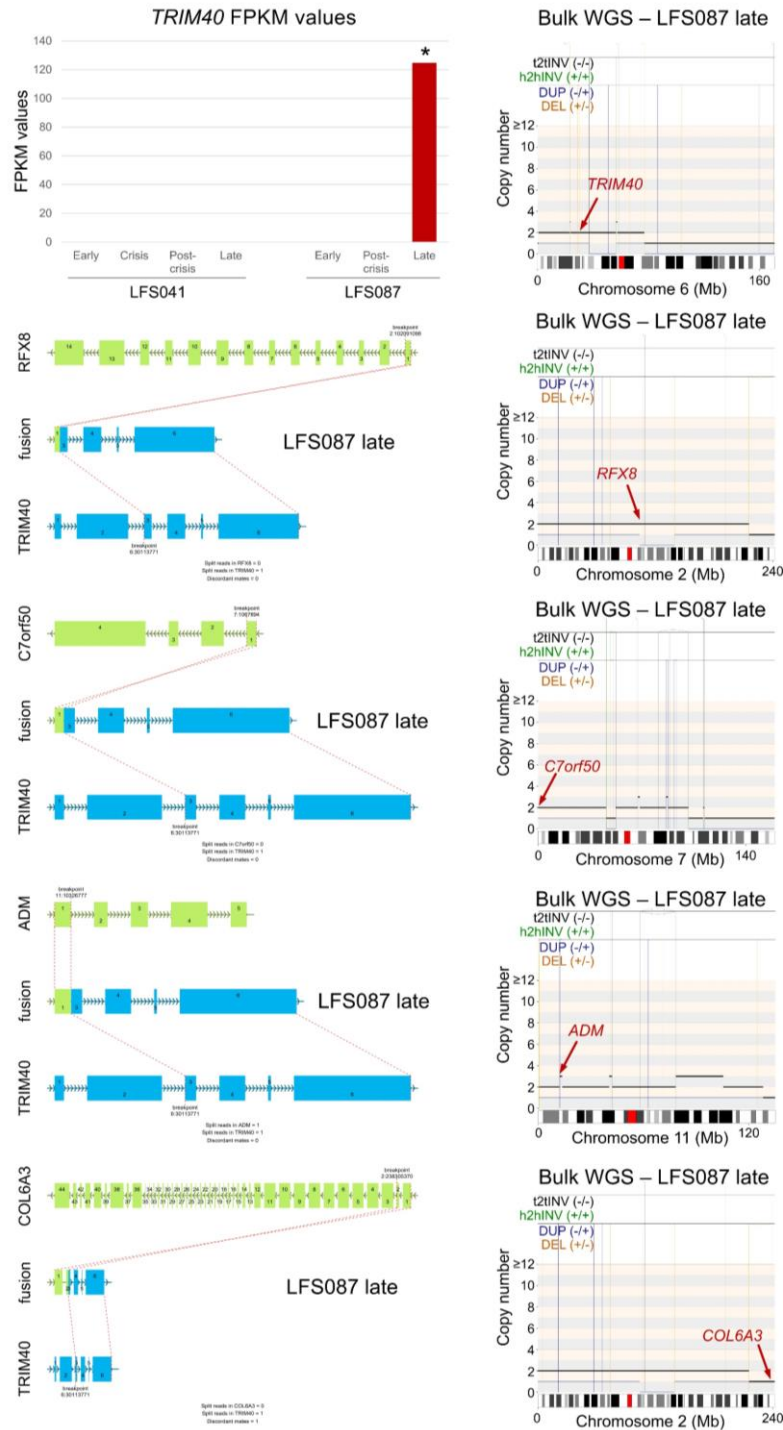
Supplementary figures



Supplementary Figure 8. Log10 of cumulative population doubling levels (PDLs) in LFS172. The graph shows the growth of *hTERT*⁺ cells (grey), along with pMX-GFP control (blue) and empty vector control (orange) cells. Graphpad prism 8 was used to create the growth curve graphs. Retroviral transduction was performed by Florian Selt.

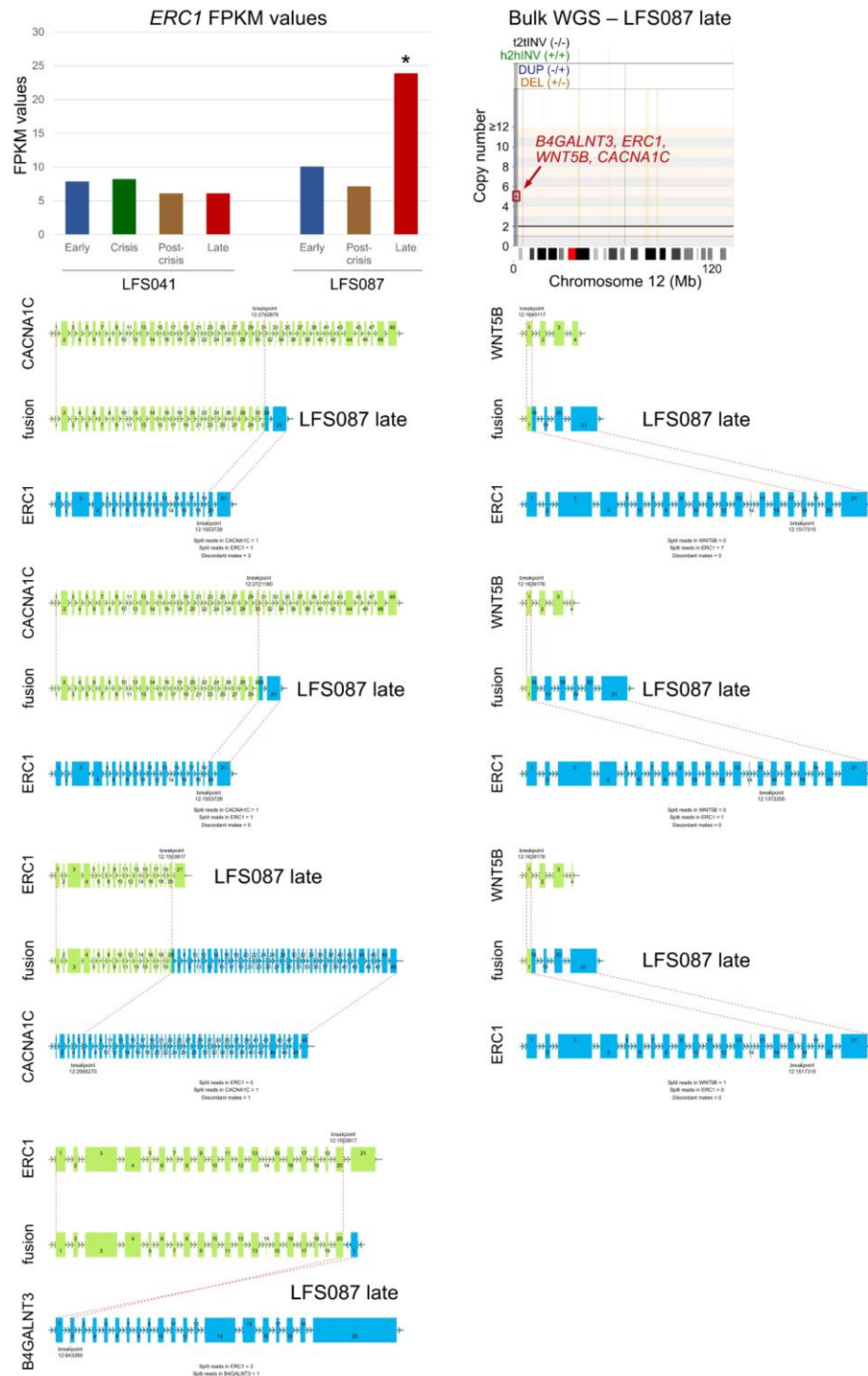


Supplementary Figure 9. ecDNA structures identified in LFS041 (p.63) and LFS087 (p.195) using AmpliconArchitect analysis of bulk WGS data. In LFS041 (p.63), amplicon 1 originating from chromosome 8 harboured the *MYC* oncogene and the *PVT1* gene. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). Michelle Chan performed the bioinformatics analysis.

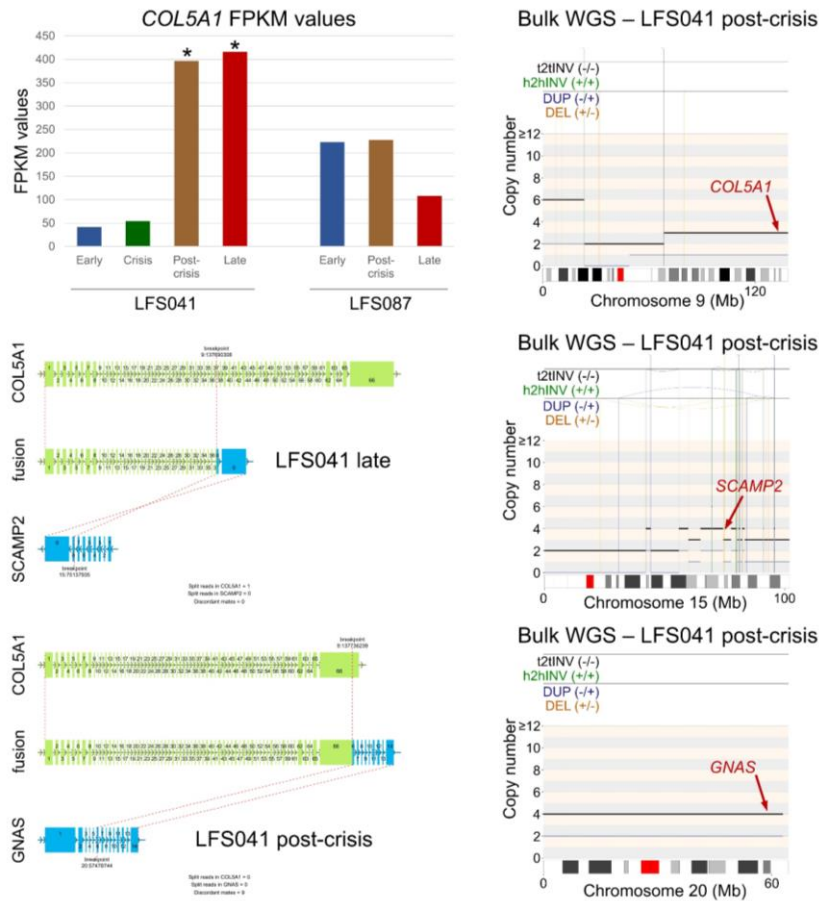


Supplementary Figure 10. TRIM40 fusions and expression in LFS primary fibroblasts. The bar plot displays FPKM values for *TRIM40* across different passages of LFS041 and LFS087 cells. The asterisk indicates the patient and the passage where the fusion was detected. Arriba plots illustrate the splice site of each gene fusion, along with the corresponding patient and passage. These include *RFX8-TRIM40*, *C7orf50-TRIM40*, *ADM-TRIM40* and *COL6A3-TRIM40*, all detected in LFS087 p.195 (late) by bulk RNA-seq. Noting that an additional *RFX8-TRIM40* gene fusion is shown in Figure 29. ReConPlots (copy number plots) from bulk WGS of LFS087 p.195 (late passage) show the gene locus on the corresponding chromosome, along with the different types of SVs. Colour coding indicates the sites of deletions (orange), duplications (blue), head-to-head inversions (green) and tail-to-tail inversions (black). Total chromosome length is shown in megabase (Mb). t2tINV: tail-to-tail inversion, h2hINV: head-to-head inversion, DUP: duplication and DEL: deletion. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). ShatterSeek analysis was performed by Petr Smirnov.

Supplementary figures



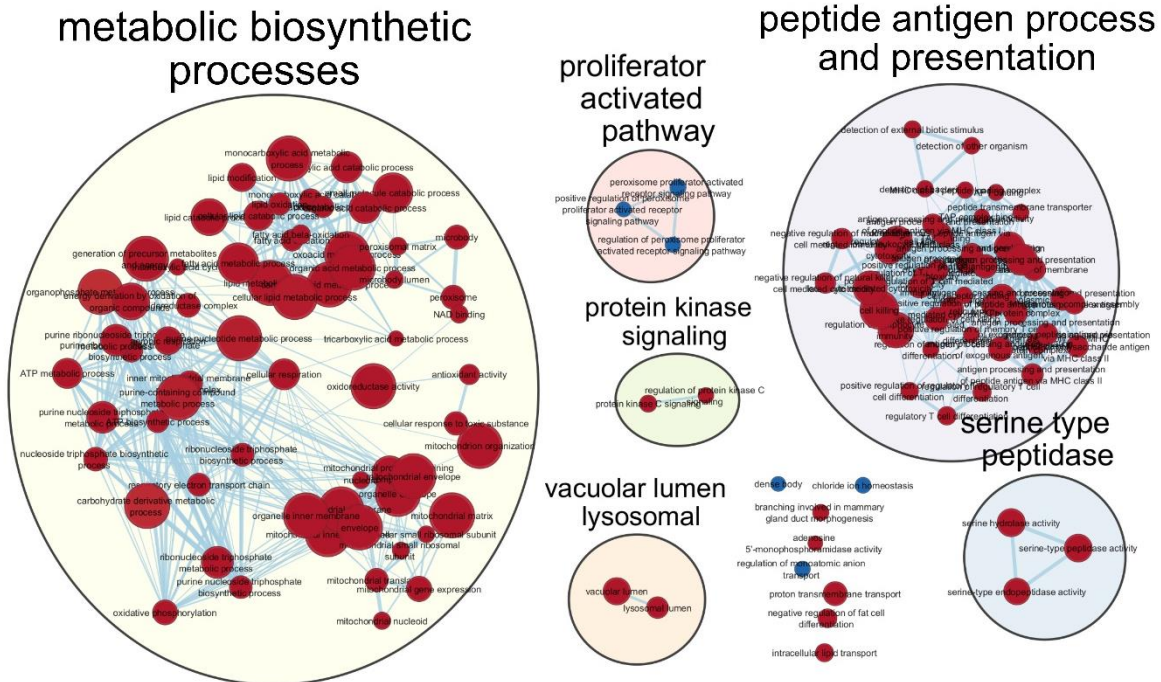
Supplementary Figure 11. ERC1 fusions and expression in LFS primary fibroblasts. The bar plot displays FPKM values for *ERC1* across different passages of LFS041 and LFS087 cells. The asterisk indicates the patient and the passage where the fusion was detected. Arriba plots illustrate the splice site of each gene fusion, along with the corresponding patient and passage. These include *CACNA1C-ERC1*, *WNT5B-ERC1* and *ERC1-B4GALNT3*, all detected in LFS087 p.195 (late) by bulk RNA-seq. ReConPlot (copy number plot) from bulk WGS of LFS087 p.195 (late passage) show the gene locus on the corresponding chromosome, along with the different types of SVs. Colour coding indicates the sites of deletions (orange), duplications (blue), head-to-head inversions (green) and tail-to-tail inversions (black). Total chromosome length is shown in megabase (Mb). t2tINV: tail-to-tail inversion, h2hINV: head-to-head inversion, DUP: duplication and DEL: deletion. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). ShatterSeek analysis was performed by Petr Smirnov.



Supplementary Figure 12. COL5A1 fusions and expression in LFS primary fibroblasts. The bar plot displays FPKM values for *COL5A1* across different passages of LFS041 and LFS087 cells, one of the fusion partners in the *COL5A1*-*SCAMP2* fusion (detected in LFS041 p.346 – late passage) and in the *COL5A1*-*GNAS* fusion (detected in LFS041 p.65 – post-crisis passage), identified through Arriba analysis of bulk RNA-seq. The asterisks indicate the patient and the passages where the fusion was detected. Arriba plots illustrate the splice site of each gene fusion, along with the corresponding patient and passage. ReConPlots (copy number plots) from bulk WGS of LFS041 p.63 (post-crisis passage) show the gene locus on the corresponding chromosome, along with the different types of SVs. Colour coding indicates the sites of deletions (orange), duplications (blue), head-to-head inversions (green) and tail-to-tail inversions (black). Total chromosome length is shown in megabase (Mb). t2tINV: tail-to-tail inversion, h2tINV: head-to-head inversion, DUP: duplication and DEL: deletion. Figure taken from Zaatra*, Philippos* et al., manuscript in preparation (see Publications, manuscript I). ShatterSeek analysis was performed by Petr Smirnov.

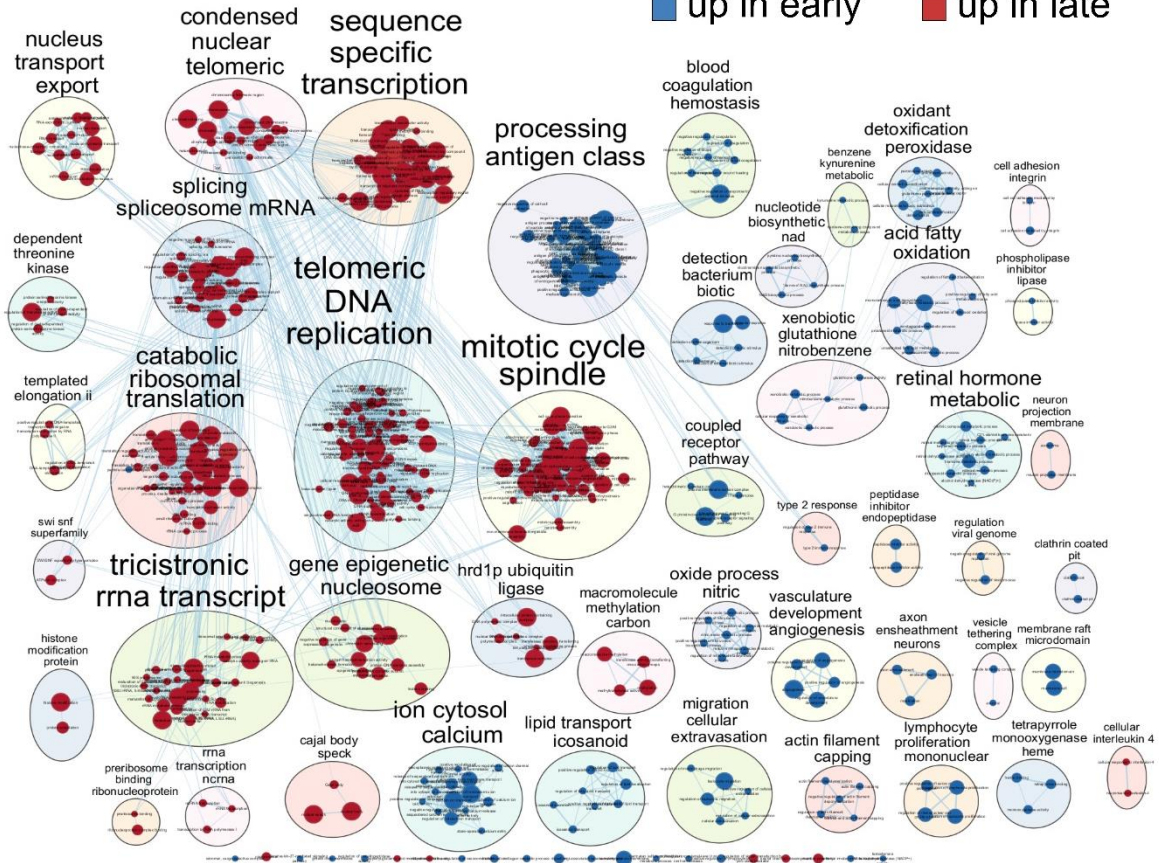
A **GSEA GO enrichment network – early vs crisis**

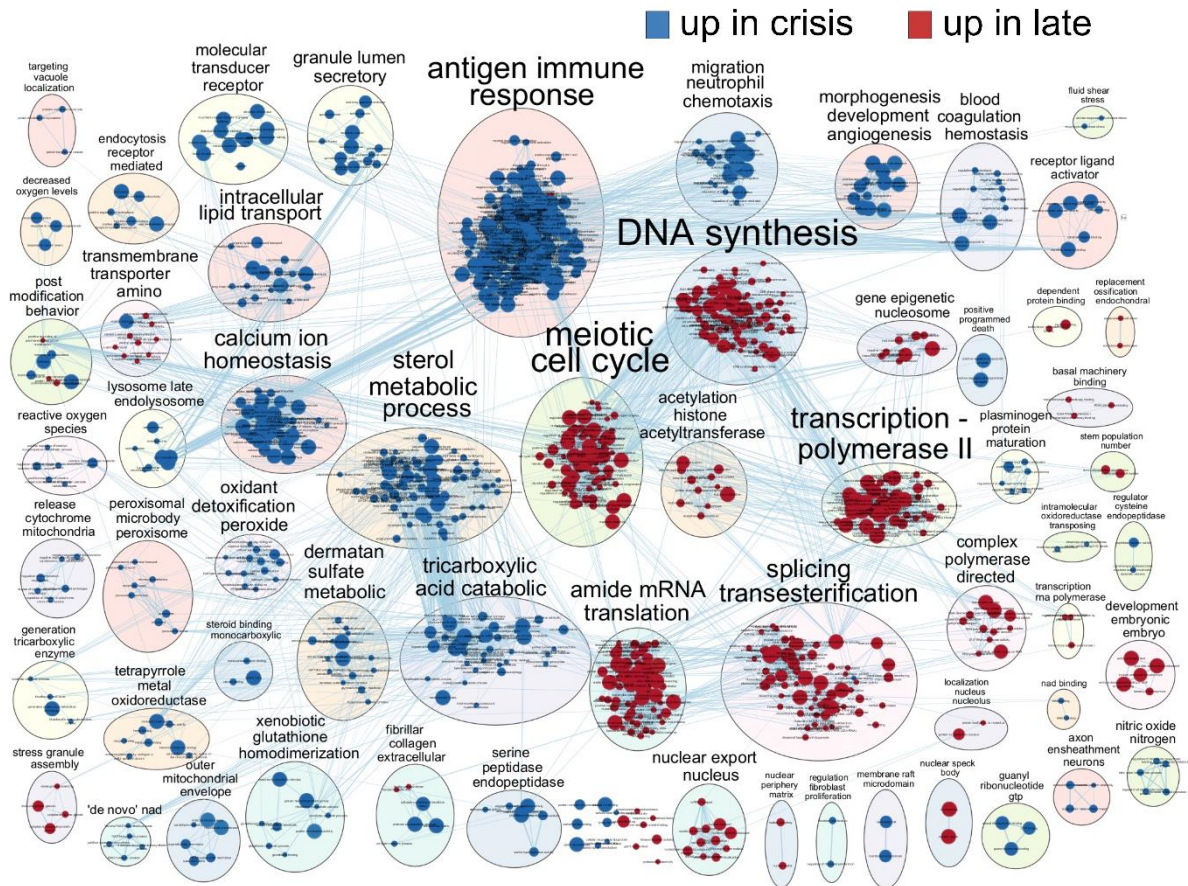
■ up in early ■ up in crisis



B **GSEA GO enrichment network – early vs late**

■ up in early ■ up in late

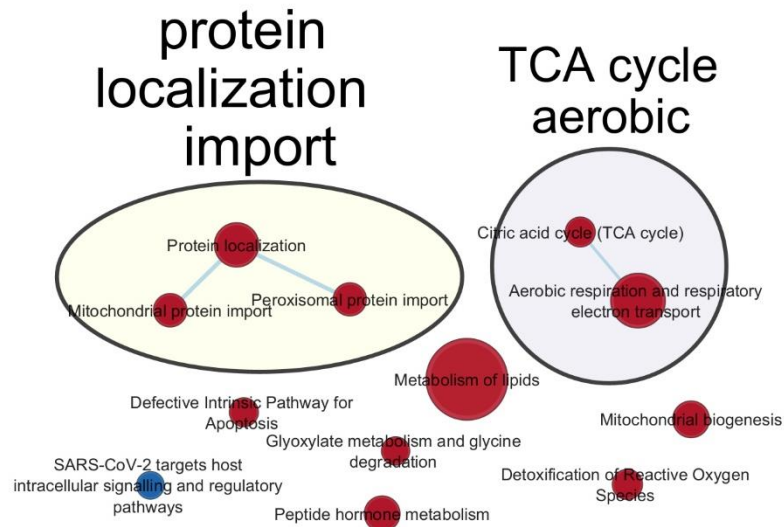


C**GSEA GO enrichment network – crisis vs late**

Supplementary Figure 13. GSEA GO enrichment network visualizing interconnected pathways in of LFS041 proteomics data. Comparisons include early (p.19) versus crisis (p.27) passages (**A**), early (p.19) versus late (p.346) passages (**B**), and crisis (p.27) versus late (p.346) passages (**C**). The network was created using the EnrichmentMap tool in Cytoscape and annotated with AutoAnnotate. Nodes represent enriched GO terms, and edges indicate the similarity between GO processes based on their shared genes.

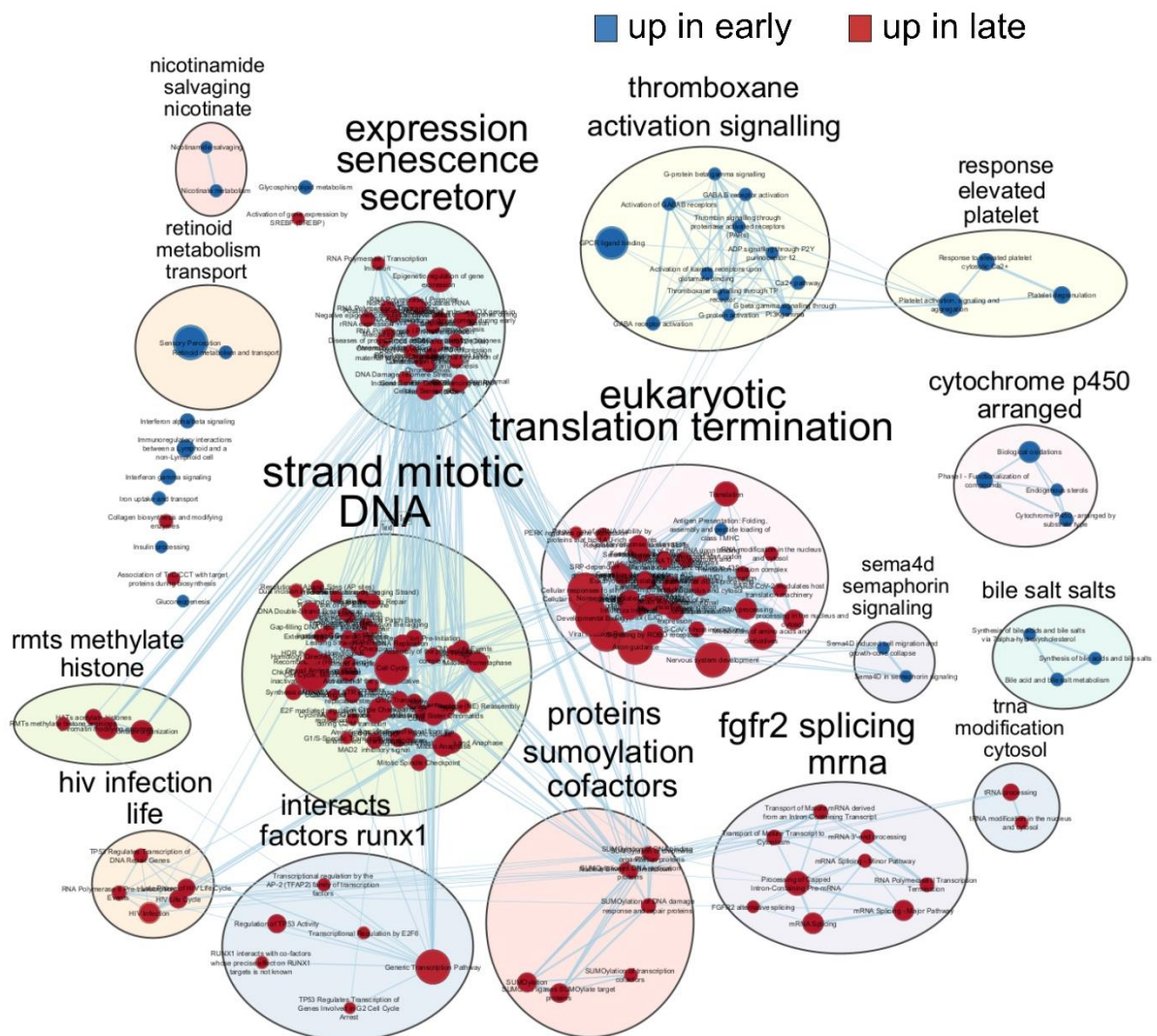
A **GSEA GO enrichment network – early vs crisis**

■ up in early ■ up in crisis



B **GSEA GO enrichment network – early vs late**

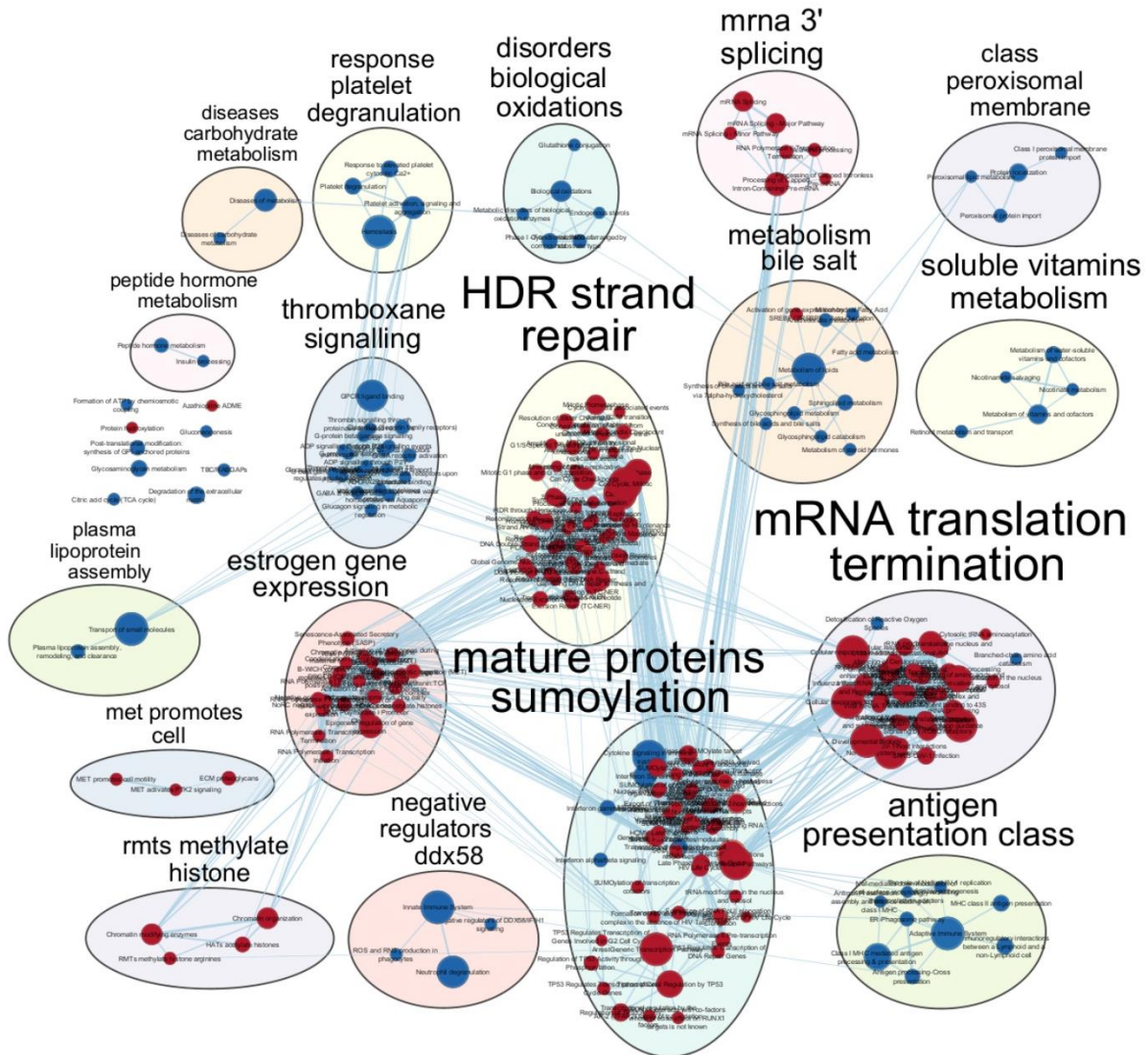
■ up in early ■ up in late



C

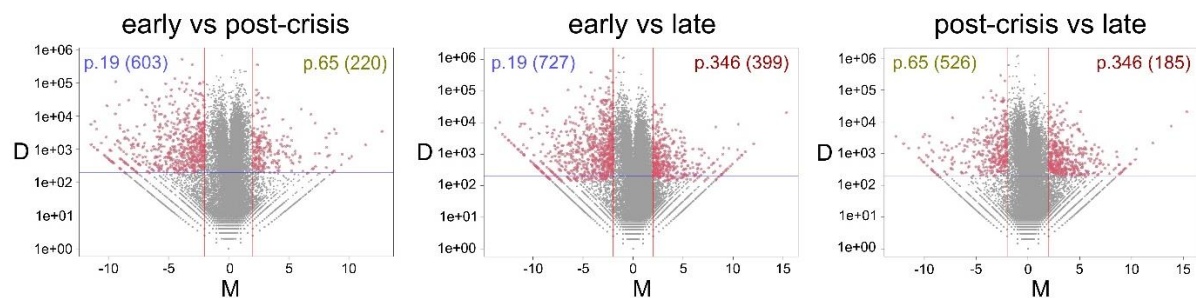
GSEA GO enrichment network – crisis vs late

■ up in crisis ■ up in late

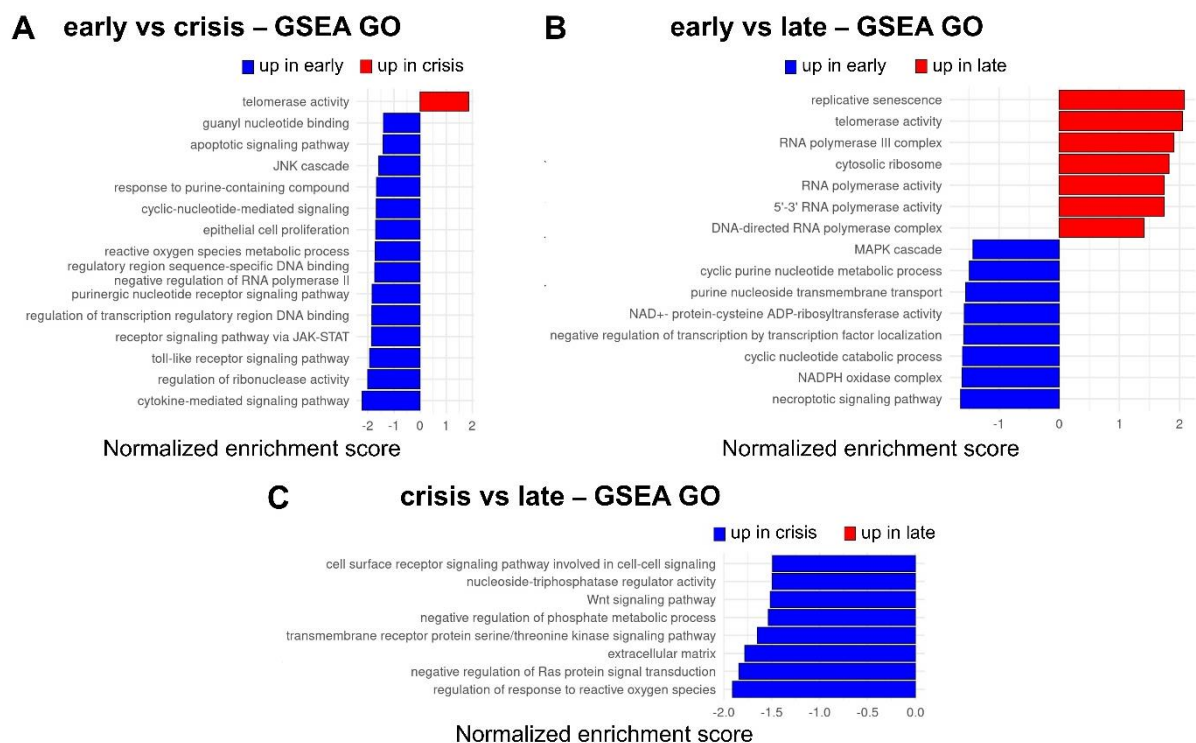


Supplementary Figure 14. GSEA Reactome enrichment network visualizing interconnected pathways in of LFS041 proteomics data. Comparisons include early (p.19) versus crisis (p.27) passages (A), early (p.19) versus late (p.346) passages (B), and crisis (p.27) versus late (p.346) passages (C). The network was created using the EnrichmentMap tool in Cytoscape and annotated with AutoAnnotate. Nodes represent enriched pathways, and edges indicate the similarity between processes and pathways based on their shared genes.

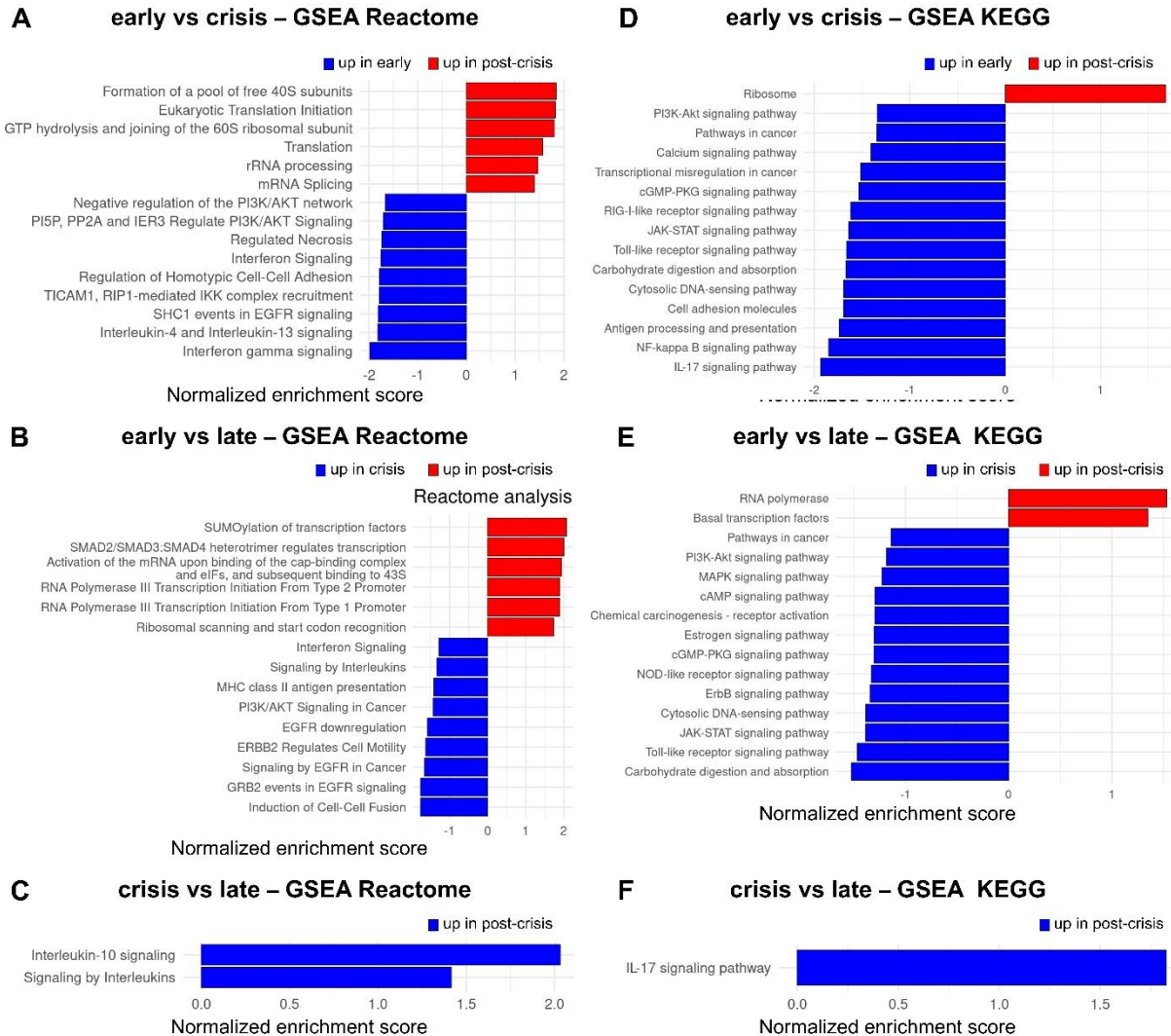
Supplementary figures



Supplementary Figure 15. Volcano plots of DEGs across cellular passages from bulk RNA-seq of LFS087. The volcano plots visualize the DEGs (highlighted with a red circle) from the LFS087 sample at three passages: early (p.20), post-crisis (p.47), and late (p.192). I performed the analysis using NOIseq. The x-axis (M) represents the log-fold change of expression levels between two conditions, indicating the magnitude and direction of change, with positive values corresponding to upregulated genes and negative values to downregulated genes. the y-axis (D) shows the absolute expression difference between conditions, reflecting the confidence that a gene is differentially expressed. number of upregulated genes at each passage is indicated in parentheses on each plot.



Supplementary Figure 16. Complementary GSEA GO analysis in LFS041 bulk RNA-seq data. Selected significantly enriched GSEA GO processes across three passages in patient LFS041. Comparisons include early (p.19) vs. post-crisis (p.65) passages (**A**), crisis (p.29) vs. post-crisis (p.65) passages (**B**), and post-crisis (p.65) vs. late (p.346) passages (**C**). The x-axis shows the normalized enrichment score, and the y-axis displays the description of the GO processes.



Supplementary Figure 17. Complementary GSEA Reactome and KEGG pathway analysis in LFS041 bulk RNA-seq data. A. B. C. Selected significantly enriched GSEA Reactome processes across three passages in patient LFS041. Comparisons include early (p.19) vs. post-crisis (p.65) passages (**A**), crisis (p.29) vs. post-crisis (p.65) passages (**B**), and post-crisis (p.65) vs. late (p.346) passages (**C**). The x-axis shows the normalized enrichment score, and the y-axis displays the description of the Reactome pathways. **D. E. F.** Selected significantly enriched GSEA KEGG processes across three passages in patient LFS041. Comparisons include early (p.19) vs. post-crisis (p.65) passages (**D**), crisis (p.29) vs. late (p.346) passages (**E**), and post-crisis (p.65) vs. late (p.346) passages (**F**). The x-axis shows the normalized enrichment score, and the y-axis displays the description of the KEGG pathways.

Publications

Publication I:

Smirnov P*, Przybilla MJ*, Simovic-Lorenz M*, Parra RG*, Susak H*, Ratnaparkhe M, Wong JK, Körber V, Mallm JP, **Philippos G**, Sill M, Kolb T, Kumar R, Casiraghi N, Okonechnikov K, Ghasemi DR, Maaß KK, Pajtler KW, Jauch A, Korshunov A, Höfer T, Zapatka M, Pfister SM, Huber W, Stegle O**, Ernst A**. Multi-omic and single-cell profiling of chromothriptic medulloblastoma reveals genomic and transcriptomic consequences of genome instability. Nat Commun. 2024 Nov 23;15(1):10183. doi: 10.1038/s41467-024-54547-w. Erratum in: Nat Commun. 2025 Jan 27;16(1):1085. doi: 10.1038/s41467-025-56164-7. PMID: 39580568; PMCID: PMC11585558.

* Shared first authors

Shared last authors and corresponding authors

Publication II:

Otoničar J*, Lazareva O*, Mallm JP*, Simovic-Lorenz M, **Philippos G**, Sant P, Parekh U, Hammann L, Li A, Yildiz U, Marttinen M, Zaugg J, Noh KM, Stegle O#, Ernst A#. HIPSD&R-seq enables scalable genomic copy number and transcriptome profiling. Genome Biol. 2024 Dec 18;25(1):316. doi: 10.1186/s13059-024-03450-0. PMID: 39696535; PMCID: PMC11657747.

* Shared first authors

Shared last authors and corresponding authors

Manuscript I:

Zaatra W*, **Philippos G***, et al., Kerem B#, Ernst A#. Nucleotide insufficiency induced by p53 deficiency leads to replication stress driving genomic instability. In preparation.

* Shared first authors

Shared last authors and corresponding authors

Acknowledgements

First of all, I would like to express my gratitude to my supervisor, PD Dr. Aurélie Ernst, for giving me the great opportunity to pursue my doctoral studies at the German Cancer Research Center (DKFZ). Thank you for your invaluable guidance, feedback, and support throughout this journey, as well as for pushing me forward, which kept my motivation going. This PhD journey not only allowed me to gain scientific knowledge and to develop a more critical and independent approach to research, but it also helped me to grow personally beyond academia and improve my soft skills, communication, problem-solving and time management.

I would like to acknowledge the contribution of my thesis advisory committee, Prof. Dr. Oliver Stegle, Prof. Dr. Peter Lichter, and Dr. Brian Luke, for their support and constructive feedback, which helped shape this work.

Additionally, I would like to thank Prof. Dr. Oliver Stegle, PD Dr. Aurélie Ernst, Prof. Dr. Dr. Georg Stöcklin and Prof. Dr. Stefan Wiemann for being part of my defence committee and for the time they took to evaluate my dissertation.

This work would not have been possible without the support and combined efforts of many people and the contributions of all collaborators. A special thanks to Michaela Hergt for the great support, especially in the beginning with imaging and staining. However, most importantly, thank you for the great help in cell culture, especially when one full incubator was not enough to fit all the plates and experimental conditions. Moreover, I am thankful to Frauke Devens for her patience and support in numerous experiments, including *TERT*-FISH, ALT-FISH, RNA extractions, as well as ploidy and ecDNA evaluation. I also wish to extend my gratitude to Michèle Vousten for all the technical support.

I am grateful for the great bioinformatics support from Dr. Petr Smirnov with bulk WGS and Strand-seq data, especially in phylogeny analysis. I also wish to thank Jan Otoničar for his analysis of scDNA sequencing data (HIPSD&R-seq and PTA). Furthermore, I would like to acknowledge Dr. Gianluca Sigismondo for performing the mass spectrometry-based proteomics and its analysis, Dr. Karen Grimes for performing Strand-seq, Dr. Florian Selt for transducing the LFS cells, and Michelle Chan for running AmpliconArchitect.

I would like to express my appreciation to the core facilities at the DKFZ for their invaluable support: Dr. Jan-Philipp Mallm, Dr. Pooja Sant, Katharina Bauer and Denise Keitel from the Single-cell Open Lab; Dr. Steffen Schmitt and Florian Blum from Flow Cytometry Core Facility; Dr. Damir Kronic from the Light Microscopy Core Facility; and to the Next Generation Sequencing Core Facility. Thank you all for the efforts and general support.

Acknowledgements

A special thank you to Prof. Batsheva Kerem and Wisam Zaatra for the fruitful collaboration and the excellent work.

I am especially grateful to all past and present members of my group (B420) for the time we spent together, the constructive feedback I always received and for all the discussions we had, both inside and outside the lab. It has been a pleasure working with all of you. I was also fortunate to be part of the B06X division. The feedback and ideas I received during my presentations were invaluable, and the seminars provided new perspectives throughout my PhD.

I am also thankful to my friends and colleagues, not only in Heidelberg or Germany but everywhere. Your support and shared moments have made this journey even more memorable.

Finally, I want to express my deepest gratitude to my family, especially my parents and my brother, who have always believed in me, supported me in every possible way, and motivated me, not only during my PhD but throughout my entire life. I am blessed to have them, and without them, I would not be where I am today.

References

1. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
2. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability--an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
3. Yao, Y. & Dai, W. Genomic Instability and Cancer. *J. Carcinog. Mutagen.* **5**, 1000165 (2014).
4. Pikor, L., Thu, K., Vucic, E. & Lam, W. The detection and implication of genome instability in cancer. *Cancer Metastasis Rev.* **32**, 341–352 (2013).
5. Loeb, L. A. A Mutator Phenotype in Cancer¹. *Cancer Res.* **61**, 3230–3239 (2001).
6. Kinzler, K. W. & Vogelstein, B. Gatekeepers and caretakers. *Nature* **386**, 761–763 (1997).
7. Rao, C. V., Yamada, H. Y., Yao, Y. & Dai, W. Enhanced genomic instabilities caused by deregulated microtubule dynamics and chromosome segregation: a perspective from genetic studies in mice. *Carcinogenesis* **30**, 1469–1474 (2009).
8. Artandi, S. E. & DePinho, R. A. Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9–18 (2010).
9. Artandi, S. E. & DePinho, R. A. Mice without telomerase: what can they teach us about human cancer? *Nat. Med.* **6**, 852–855 (2000).
10. The two sides of chromosomal instability: drivers and brakes in cancer | Signal Transduction and Targeted Therapy. <https://www.nature.com/articles/s41392-024-01767-7>.
11. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
12. Saxena, S. & Zou, L. Hallmarks of DNA replication stress. *Mol. Cell* **82**, 2298–2314 (2022).
13. Goehring, L., Huang, T. T. & Smith, D. J. Transcription-Replication Conflicts as a Source of Genome Instability. *Annu. Rev. Genet.* **57**, 157–179 (2023).
14. Langie, S. A. S. *et al.* Causes of genome instability: the effect of low dose chemical exposures in modern society. *Carcinogenesis* **36**, S61–S88 (2015).

References

15. Requesens, M., Fojer, F., Nijman, H. W. & de Bruyn, M. Genomic instability as a driver and suppressor of anti-tumor immunity. *Front. Immunol.* **15**, 1462496 (2024).
16. Chen, M., Linstra, R. & van Vugt, M. A. T. M. Genomic instability, inflammatory signaling and response to cancer immunotherapy. *Biochim. Biophys. Acta Rev. Cancer* **1877**, 188661 (2022).
17. Cahill, D. P., Kinzler, K. W., Vogelstein, B. & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol.* **9**, M57-60 (1999).
18. Podlaha, O., Riester, M., De, S. & Michor, F. Evolution of the cancer genome. *Trends Genet.* **28**, 155–163 (2012).
19. Stephens, P. J. *et al.* Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* **144**, 27–40 (2011).
20. Simovic-Lorenz, M. & Ernst, A. Chromothripsis in cancer. *Nat. Rev. Cancer* (2024) doi:10.1038/s41568-024-00769-5.
21. Mazzagatti, A., Engel, J. L. & Ly, P. Boveri and beyond: Chromothripsis and genomic instability from mitotic errors. *Mol. Cell* **84**, 55–69 (2024).
22. Davis, A., Gao, R. & Navin, N. Tumor evolution: Linear, branching, neutral or punctuated? *Biochim. Biophys. Acta BBA - Rev. Cancer* **1867**, 151–161 (2017).
23. Cortés-Ciriano, I. *et al.* Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat. Genet.* **52**, 331–341 (2020).
24. Voronina, N. *et al.* The landscape of chromothripsis across adult cancer types. *Nat. Commun.* **11**, 2320 (2020).
25. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
26. Fontana, M. C. *et al.* Chromothripsis in Acute Myeloid Leukemia: biological features and impact on survival. *Leukemia* 10.1038/leu.2017.351 (2017) doi:10.1038/leu.2017.351.
27. Bolkestein, M. *et al.* Chromothripsis in Human Breast Cancer. *Cancer Res.* **80**, 4918–4931 (2020).

28. Behjati, S. *et al.* Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat. Commun.* **8**, 15936 (2017).
29. Ejaz, U. *et al.* Chromothripsis: an emerging crossroad from aberrant mitosis to therapeutic opportunities. *J. Mol. Cell Biol.* **16**, mjae016 (2024).
30. Koltsova, A. S. *et al.* On the Complexity of Mechanisms and Consequences of Chromothripsis: An Update. *Front. Genet.* **10**, 393 (2019).
31. Shorokhova, M., Nikolsky, N. & Grinchuk, T. Chromothripsis—Explosion in Genetic Science. *Cells* **10**, 1102 (2021).
32. Rode, A., Maass, K. K., Willmund, K. V., Lichter, P. & Ernst, A. Chromothripsis in cancer cells: An update. *Int. J. Cancer* **138**, 2322–2333 (2016).
33. Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148**, 59–71 (2012).
34. Oliner, J. D., Saiki, A. Y. & Caenepeel, S. The Role of MDM2 Amplification and Overexpression in Tumorigenesis. *Cold Spring Harb. Perspect. Med.* **6**, a026336 (2016).
35. Onel, K. & Cordon-Cardo, C. MDM2 and Prognosis. *Mol. Cancer Res.* **2**, 1–8 (2004).
36. Ratnaparkhe, M. *et al.* Genomic profiling of Acute lymphoblastic leukemia in ataxia telangiectasia patients reveals tight link between ATM mutations and chromothripsis. *Leukemia* **31**, 2048–2056 (2017).
37. Ostapińska, K., Styka, B. & Lejman, M. Insight into the Molecular Basis Underlying Chromothripsis. *Int. J. Mol. Sci.* **23**, 3318 (2022).
38. Pellestor, F., Gaillard, J. B., Schneider, A., Puechberty, J. & Gatinois, V. Chromoanagenesis, the mechanisms of a genomic chaos. *Semin. Cell Dev. Biol.* **123**, 90–99 (2022).
39. Anzick, S. *et al.* Chromoanasythesis as a cause of Jacobsen syndrome. *Am. J. Med. Genet. A.* **182**, 2533–2539 (2020).
40. Lin, Y.-F. *et al.* Mitotic clustering of pulverized chromosomes from micronuclei. *Nature* **618**, 1041–1048 (2023).

References

41. Maher, C. A. & Wilson, R. K. Chromothripsis and human disease: piecing together the shattering process. *Cell* **148**, 29–32 (2012).
42. Espejo Valle-Inclan, J. *et al.* Ongoing chromothripsis underpins osteosarcoma genome complexity and clonal evolution. *Cell* **188**, 352–370.e22 (2025).
43. McDermott, D. H. *et al.* Chromothriptic cure of WHIM syndrome. *Cell* **160**, 686–699 (2015).
44. Korbel, J. O. & Campbell, P. J. Criteria for Inference of Chromothripsis in Cancer Genomes. *Cell* **152**, 1226–1236 (2013).
45. Chen, C.-Y. Chromothripsis in myeloid malignancies. *Ann. Hematol.* **103**, 3955–3962 (2024).
46. Yang, J. *et al.* CTLPScanner: a web server for chromothripsis-like pattern detection. *Nucleic Acids Res.* **44**, W252–258 (2016).
47. Govind, S. K. *et al.* ShatterProof: operational detection and quantification of chromothripsis. *BMC Bioinformatics* **15**, 78 (2014).
48. Balajee, A. S. & Hande, M. P. History and evolution of cytogenetic techniques: Current and future applications in basic and clinical research. *Mutat. Res. Toxicol. Environ. Mutagen.* **836**, 3–12 (2018).
49. Ernst, A. *et al.* Telomere dysfunction and chromothripsis. *Int. J. Cancer* **138**, 2905–2914 (2016).
50. Jia, Q., Chu, H., Jin, Z., Long, H. & Zhu, B. High-throughput single-cell sequencing in cancer research. *Signal Transduct. Target. Ther.* **7**, 1–20 (2022).
51. Smirnov, P. *et al.* Multi-omic and single-cell profiling of chromothriptic medulloblastoma reveals genomic and transcriptomic consequences of genome instability. *Nat. Commun.* **15**, 10183 (2024).
52. Otoničar, J. *et al.* HIPSD&R-seq enables scalable genomic copy number and transcriptome profiling. *Genome Biol.* **25**, 316 (2024).
53. Falconer, E. *et al.* DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. *Nat. Methods* **9**, 1107–1112 (2012).

54. Sanders, A. D., Falconer, E., Hills, M., Spierings, D. C. J. & Lansdorp, P. M. Single-cell template strand sequencing by Strand-seq enables the characterization of individual homologs. *Nat. Protoc.* **12**, 1151–1176 (2017).
55. Sanders, A. D. *et al.* Single-cell analysis of structural variations and complex rearrangements with tri-channel processing. *Nat. Biotechnol.* **38**, 343–354 (2020).
56. Cretu Stancu, M. *et al.* Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat. Commun.* **8**, 1326 (2017).
57. Rausch, T. *et al.* Long-read sequencing of diagnosis and post-therapy medulloblastoma reveals complex rearrangement patterns and epigenetic signatures. *Cell Genomics* **3**, 100281 (2023).
58. Kats, I. *et al.* Spatio-temporal transcriptomics of chromothriptic SHH-medulloblastoma identifies multiple genetic clones that resist treatment and drive relapse. *Nat. Commun.* **15**, 10370 (2024).
59. Mammel, A. E. & Hatch, E. M. Genome instability from nuclear catastrophe and DNA damage. *Semin. Cell Dev. Biol.* **123**, 131–139 (2022).
60. Dewhurst, S. M. *et al.* Structural variant evolution after telomere crisis. *Nat. Commun.* **12**, 2093 (2021).
61. von Morgen, P. & Maciejowski, J. The ins and outs of telomere crisis in cancer. *Genome Med.* **10**, 89 (2018).
62. Bhargava, R., Fischer, M. & O’Sullivan, R. J. Genome rearrangements associated with aberrant telomere maintenance. *Curr. Opin. Genet. Dev.* **60**, 31–40 (2020).
63. Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* **163**, 1641–1654 (2015).
64. Maciejowski, J. *et al.* APOBEC3-dependent kataegis and TREX1-driven chromothripsis during telomere crisis. *Nat. Genet.* **52**, 884–890 (2020).
65. Umbreit, N. T. *et al.* Mechanisms generating cancer genome complexity from a single cell division error. *Science* **368**, eaba0712 (2020).

References

66. Ly, P. *et al.* Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nat. Genet.* **51**, 705–715 (2019).
67. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
68. Crasta, K. *et al.* DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**, 53–58 (2012).
69. Hatch, E. M. & Hetzer, M. W. Nuclear envelope rupture is induced by actin-based nucleus confinement. *J. Cell Biol.* **215**, 27–36 (2016).
70. Krupina, K., Goginashvili, A. & Cleveland, D. W. Causes and consequences of micronuclei. *Curr. Opin. Cell Biol.* **70**, 91–99 (2021).
71. Hatch, E. M., Fischer, A. H., Deerinck, T. J. & Hetzer, M. W. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell* **154**, 47–60 (2013).
72. Mohr, L. *et al.* ER-directed TREX1 limits cGAS activation at micronuclei. *Mol. Cell* **81**, 724–738.e9 (2021).
73. Kwon, M., Leibowitz, M. L. & Lee, J.-H. Small but mighty: the causes and consequences of micronucleus rupture. *Exp. Mol. Med.* **52**, 1777–1786 (2020).
74. Zhang, C.-Z. *et al.* Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
75. Petsalaki, E. & Zachos, G. The Abscission Checkpoint: A Guardian of Chromosomal Stability. *Cells* **10**, 3350 (2021).
76. Mocanu, C. & Chan, K.-L. Mind the replication gap. *R. Soc. Open Sci.* **8**, 201932 (2021).
77. Gelot, C., Magdalou, I. & Lopez, B. S. Replication stress in Mammalian cells and its consequences for mitosis. *Genes* **6**, 267–298 (2015).
78. Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nat. Cell Biol.* **16**, 2–9 (2014).
79. Kotsantis, P., Petermann, E. & Boulton, S. J. Mechanisms of Oncogene-Induced Replication Stress: Jigsaw Falling into Place. *Cancer Discov.* **8**, 537–555 (2018).

80. Shaikh, N. *et al.* Replication stress generates distinctive landscapes of DNA copy number alterations and chromosome scale losses. *Genome Biol.* **23**, 223 (2022).
81. Huang, M. *et al.* Cell cycle progression of under-replicated cells. *Nucleic Acids Res.* **53**, gkae1311 (2025).
82. Hills, S. A. & Diffley, J. F. X. DNA replication and oncogene-induced replicative stress. *Curr. Biol. CB* **24**, R435-444 (2014).
83. Wilhelm, T. *et al.* Mild replication stress causes chromosome mis-segregation via premature centriole disengagement. *Nat. Commun.* **10**, 3585 (2019).
84. Ratnaparkhe, M. *et al.* Defective DNA damage repair leads to frequent catastrophic genomic events in murine and human tumors. *Nat. Commun.* **9**, 4760 (2018).
85. Northcott, P. A. *et al.* Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**, 49–56 (2012).
86. Pastorczak, A. *et al.* Genetic hallmarks and clinical implications of chromothripsis in childhood T-cell acute lymphoblastic leukemia. *Leukemia* **38**, 2344–2354 (2024).
87. Forment, J. V., Kaidi, A. & Jackson, S. P. Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nat. Rev. Cancer* **12**, 663–670 (2012).
88. Chen, Y., Qiu, Q., She, J. & Yu, J. Extrachromosomal circular DNA in colorectal cancer: biogenesis, function and potential as therapeutic target. *Oncogene* **42**, 941–951 (2023).
89. Wang, Y., Huang, R., Zheng, G. & Shen, J. Small ring has big potential: insights into extrachromosomal DNA in cancer. *Cancer Cell Int.* **21**, 236 (2021).
90. Dong, Y. *et al.* Extrachromosomal DNA (ecDNA) in cancer: mechanisms, functions, and clinical implications. *Front. Oncol.* **13**, 1194405 (2023).
91. Tang, J. *et al.* Enhancing transcription-replication conflict targets ecDNA-positive cancers. *Nature* **635**, 210–218 (2024).
92. Luebeck, J. *et al.* Extrachromosomal DNA in the cancerous transformation of Barrett's oesophagus. *Nature* **616**, 798–805 (2023).

References

93. Pecorino, L. T., Verhaak, R. G. W., Henssen, A. & Mischel, P. S. Extrachromosomal DNA (ecDNA): an origin of tumor heterogeneity, genomic remodeling, and drug resistance. *Biochem. Soc. Trans.* **50**, 1911–1920 (2022).
94. Molenaar, J. J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483**, 589–593 (2012).
95. Paolini, L., Hussain, S. & Galardy, P. J. Chromosome instability in neuroblastoma: A pathway to aggressive disease. *Front. Oncol.* **12**, (2022).
96. Simovic, M. *et al.* Carbon ion radiotherapy eradicates medulloblastomas with chromothripsis in an orthotopic Li-Fraumeni patient-derived mouse model. *Neuro-Oncol.* **23**, 2028–2041 (2021).
97. Hirsch, D. *et al.* Chromothripsis and focal copy number alterations determine poor outcome in malignant melanoma. *Cancer Res.* **73**, 1454–1460 (2013).
98. Shoshani, O. *et al.* Chromothripsis drives the evolution of gene amplification in cancer. *Nature* **591**, 137–141 (2021).
99. Helleday, T. Putting poly (ADP-ribose) polymerase and other DNA repair inhibitors into clinical practice. *Curr. Opin. Oncol.* **25**, 609–614 (2013).
100. Powell, S. N. & Kachnic, L. A. Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation. *Oncogene* **22**, 5784–5791 (2003).
101. Khalid, U. *et al.* A synergistic interaction between HDAC- and PARP inhibitors in childhood tumors with chromothripsis. *Int. J. Cancer* **151**, 590–606 (2022).
102. Chen, X., Agustinus, A. S., Li, J., DiBona, M. & Bakhoun, S. F. Chromosomal instability as a driver of cancer progression. *Nat. Rev. Genet.* **26**, 31–46 (2025).
103. Siri, S. O., Martino, J. & Gottifredi, V. Structural Chromosome Instability: Types, Origins, Consequences, and Therapeutic Opportunities. *Cancers* **13**, 3056 (2021).
104. Ly, P. *et al.* Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by canonical NHEJ. *Nat. Cell Biol.* **19**, 68–75 (2017).

105. Nielsen, C. F. *et al.* PICH promotes sister chromatid disjunction and co-operates with topoisomerase II in mitosis. *Nat. Commun.* **6**, 8962 (2015).
106. Leibowitz, M. L. *et al.* Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat. Genet.* **53**, 895–905 (2021).
107. Mardin, B. R. *et al.* A cell-based model system links chromothripsis with hyperploidy. *Mol. Syst. Biol.* **11**, 828 (2015).
108. Zhao, M. *et al.* CGAS is a micronucleophagy receptor for the clearance of micronuclei. *Autophagy* **17**, 3976–3991 (2021).
109. Li, J. S. Z. *et al.* Chromosomal fragile site breakage by EBV-encoded EBNA1 at clustered repeats. *Nature* **616**, 504–509 (2023).
110. Dacus, D. *et al.* Beta Human Papillomavirus 8 E6 Induces Micronucleus Formation and Promotes Chromothripsis. *J. Virol.* **96**, e01015-22.
111. Schütze, D. M. *et al.* immortalization capacity of HPV types is inversely related to chromosomal instability. *Oncotarget* **7**, 37608–37621 (2016).
112. Kolb, T. *et al.* A versatile system to introduce clusters of genomic double-strand breaks in large cell populations. *Genes. Chromosomes Cancer* **60**, 303–313 (2021).
113. Morishita, M. *et al.* Chromothripsis-like chromosomal rearrangements induced by ionizing radiation using proton microbeam irradiation system. *Oncotarget* **7**, 10182–10192 (2016).
114. Tubio, J. M. C. & Estivill, X. When catastrophe strikes a cell. *Nature* **470**, 476–477 (2011).
115. Bolkestein, M. *et al.* Chromothripsis in Human Breast Cancer. *Cancer Res.* **80**, 4918–4931 (2020).
116. Arniani, S. *et al.* Chromothripsis is a frequent event and underlies typical genetic changes in early T-cell precursor lymphoblastic leukemia in adults. *Leukemia* **36**, 2577–2585 (2022).
117. Yu, J. *et al.* Chromothripsis detection with multiple myeloma patients based on deep graph learning. *Bioinforma. Oxf. Engl.* **39**, btad422 (2023).

References

118. Schneider, K., Zelle, K., Nichols, K. E., Schwartz Levine, A. & Garber, J. Li-Fraumeni Syndrome. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).
119. Li, F. P. & Fraumeni, J. F. Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome. *J. Natl. Cancer Inst.* **43**, 1365–1373 (1969).
120. Bischoff, F. Z. *et al.* Spontaneous Abnormalities in Normal Fibroblasts from Patients with Li-Fraumeni Cancer Syndrome: Aneuploidy and Immortalization. *Cancer Res.* **50**, 7979–7984 (1990).
121. Gollahon, L. S. *et al.* Telomerase activity during spontaneous immortalization of Li-Fraumeni syndrome skin fibroblasts. (1988).
122. Little, J. B. *et al.* Normal cytotoxic response of skin fibroblasts from patients with Li-Fraumeni familial cancer syndrome to DNA-damaging agents in vitro. *Cancer Res.* **47**, 4229–4234 (1987).
123. Chen, S. Z., Kazim, D., Kravets, J. & Pollack, R. E. Skin fibroblasts from individuals hemizygous for the familial adenomatous polyposis susceptibility gene show delayed crisis in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2008–2012 (1989).
124. Wright, W. E. & Shay, J. W. The two-stage mechanism controlling cellular senescence and immortalization. *Exp. Gerontol.* **27**, 383–389 (1992).
125. Counter, C. M. *et al.* Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* **11**, 1921–1929 (1992).
126. Wright, W. E., Pereira-Smith, O. M. & Shay, J. W. Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts. *Mol. Cell. Biol.* **9**, 3088–3092 (1989).
127. Liu, P. K., Kraus, E., Wu, T. A., Strong, L. C. & Tainsky, M. A. Analysis of genomic instability in Li-Fraumeni fibroblasts with germline p53 mutations. *Oncogene* **12**, 2267–2278 (1996).

128. Li, Y. *et al.* Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* **508**, 98–102 (2014).
129. Ly, P. & Cleveland, D. W. Rebuilding Chromosomes After Catastrophe: Emerging Mechanisms of Chromothripsis. *Trends Cell Biol.* **27**, 917–930 (2017).
130. Boisselier, B. *et al.* Whole genome duplication is an early event leading to aneuploidy in IDH -wild type glioblastoma. *Oncotarget* **9**, 36017–36028 (2018).
131. Prickett, T. D. & Samuels, Y. Molecular pathways: dysregulated glutamatergic signaling pathways in cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **18**, 4240–4246 (2012).
132. Zhou, X. J., Lin, B. R. & Wang, Y. Expression of GRIN2A in benign and malignant nasopharyngeal diseases and its clinical significance. *Genet. Mol. Res. GMR* **14**, 17289–17295 (2015).
133. Prickett, T. D. *et al.* Somatic mutation of GRIN2A in malignant melanoma results in loss of tumor suppressor activity via aberrant NMDAR complex formation. *J. Invest. Dermatol.* **134**, 2390–2398 (2014).
134. D'mello, S. A. N. *et al.* Evidence That GRIN2A Mutations in Melanoma Correlate with Decreased Survival. *Front. Oncol.* **3**, 333 (2014).
135. Tsutsui, T. *et al.* Immortal, telomerase-negative cell lines derived from a Li-Fraumeni syndrome patient exhibit telomere length variability and chromosomal and minisatellite instabilities. *Carcinogenesis* **24**, 953–965 (2003).
136. Gonzalez-Pena, V. *et al.* Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc. Natl. Acad. Sci.* **118**, e2024176118 (2021).
137. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLOS ONE* **5**, e13984 (2010).
138. Kucera, M., Isserlin, R., Arkhangorodsky, A. & Bader, G. D. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Research* **5**, 1717 (2016).

References

139. Stincone, A. *et al.* The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biol. Rev. Camb. Philos. Soc.* **90**, 927–963 (2015).
140. Keenan, A. B. *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res.* **47**, W212–W224 (2019).
141. Nguyen, D. Q. *et al.* Ebp1 p48 promotes oncogenic activities in human colon cancer cells through regulation of TIF-90-mediated ribosomal RNA synthesis. *J. Cell. Physiol.* **234**, 17612–17621 (2019).
142. Wild, K. *et al.* MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. *Nat. Commun.* **11**, 776 (2020).
143. Liang, X. *et al.* Structural snapshots of human pre-60S ribosomal particles before and after nuclear export. *Nat. Commun.* **11**, 3542 (2020).
144. Stevenson, B. W. *et al.* A structural view of PA2G4 isoforms with opposing functions in cancer. *J. Biol. Chem.* **295**, 16100–16112 (2020).
145. Sun, S. *et al.* PA2G4 promotes the metastasis of hepatocellular carcinoma by stabilizing FYN mRNA in a YTHDF2-dependent manner. *Cell Biosci.* **12**, 55 (2022).
146. Bachand, F. & Silver, P. A. PRMT3 is a ribosomal protein methyltransferase that affects the cellular levels of ribosomal subunits. *EMBO J.* **23**, 2641–2650 (2004).
147. Swiercz, R., Person, M. D. & Bedford, M. T. Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). *Biochem. J.* **386**, 85–91 (2005).
148. Swiercz, R., Cheng, D., Kim, D. & Bedford, M. T. Ribosomal Protein rpS2 Is Hypomethylated in PRMT3-deficient Mice *. *J. Biol. Chem.* **282**, 16917–16923 (2007).
149. Dionne, K. L., Bergeron, D., Landry-Voyer, A.-M. & Bachand, F. The 40S ribosomal protein uS5 (RPS2) assembles into an extraribosomal complex with human ZNF277 that competes with the PRMT3–uS5 interaction. *J. Biol. Chem.* **294**, 1944–1955 (2019).
150. Lei, Y. *et al.* Protein arginine methyltransferase 3 promotes glycolysis and hepatocellular carcinoma growth by enhancing arginine methylation of lactate dehydrogenase A. *Clin. Transl. Med.* **12**, e686 (2022).

151. Hsu, M.-C. *et al.* Protein arginine methyltransferase 3-induced metabolic reprogramming is a vulnerable target of pancreatic cancer. *J. Hematol. Oncol.* **12**, 79 (2019).
152. Liao, Y. *et al.* PRMT3 drives glioblastoma progression by enhancing HIF1A and glycolytic metabolism. *Cell Death Dis.* **13**, 1–14 (2022).
153. Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: a matter of depth. *Genome Res.* **21**, 2213–2223 (2011).
154. Tarazona, S. *et al.* Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* **43**, e140 (2015).
155. Seyednasrollah, F., Laiho, A. & Elo, L. L. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinform.* **16**, 59–70 (2015).
156. Schurch, N. J. *et al.* How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* **22**, 839–851 (2016).
157. DNA-binding transcription activator activity Gene Ontology Term (GO:0001216). https://www.informatics.jax.org/vocab/gene_ontology/GO:0001216.
158. QuickGO::Term GO:0001216. <https://www.ebi.ac.uk/QuickGO/term/GO:0001216>.
159. Carbon, S. & Mungall, C. Gene Ontology Data Archive. Zenodo <https://doi.org/10.5281/ZENODO.14083199> (2024).
160. Talukdar, P. D. & Chatterji, U. Transcriptional co-activators: emerging roles in signaling pathways and potential therapeutic targets for diseases. *Signal Transduct. Target. Ther.* **8**, 1–41 (2023).
161. Tian, M. *et al.* Integrative analysis of DNA replication origins and ORC-/MCM-binding sites in human cells reveals a lack of overlap. *eLife* **12**, RP89548 (2024).
162. Primo, L. M. F. & Teixeira, L. K. DNA replication stress: oncogenes in the spotlight. *Genet. Mol. Biol.* **43**, e20190138 (2019).
163. Bowry, A., Kelly, R. D. W. & Petermann, E. Hypertranscription and replication stress in cancer. *Trends Cancer* **7**, 863–877 (2021).

References

164. Zatzman, M. *et al.* Widespread hypertranscription in aggressive human cancers. *Sci. Adv.* **8**, eabn0238 (2022).
165. Lin, C. Y. *et al.* Transcriptional Amplification in Tumor Cells with Elevated c-Myc. *Cell* **151**, 56–67 (2012).
166. Lynch, C. J. & Milner, J. Loss of one p53 allele results in four-fold reduction of p53 mRNA and protein: a basis for p53 haplo-insufficiency. *Oncogene* **25**, 3463–3470 (2006).
167. Venkatachalam, S. *et al.* Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation. *EMBO J.* **17**, 4657–4667 (1998).
168. Merino, D. & Malkin, D. p53 and hereditary cancer. *Subcell. Biochem.* **85**, 1–16 (2014).
169. Shay, J. W. & Wright, W. E. Quantitation of the frequency of immortalization of normal human diploid fibroblasts by SV40 large T-antigen. *Exp. Cell Res.* **184**, 109–118 (1989).
170. Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp. Cell Res.* **25**, 585–621 (1961).
171. Tsutsui, T. *et al.* Aflatoxin B1-induced immortalization of cultured skin fibroblasts from a patient with Li-Fraumeni syndrome. *Carcinogenesis* **16**, 25–34 (1995).
172. Endo, H. & Inoue, M. Dormancy in cancer. *Cancer Sci.* **110**, 474–480 (2019).
173. Aguirre-Ghiso, J. A. Models, mechanisms and clinical evidence for cancer dormancy. *Nat. Rev. Cancer* **7**, 834–846 (2007).
174. Karlsson, K. *et al.* Deterministic evolution and stringent selection during preneoplasia. *Nature* **618**, 383–393 (2023).
175. Dreesen, O., Li, B. & Cross, G. A. M. Telomere structure and shortening in telomerase-deficient *Trypanosoma brucei*. *Nucleic Acids Res.* **33**, 4536–4543 (2005).
176. Du, H.-Y. *et al.* Telomerase reverse transcriptase haploinsufficiency and telomere length in individuals with 5p– syndrome. *Aging Cell* **6**, 689–697 (2007).
177. Cheng, L., Li, J., Sun, H. & Jiang, H. Appropriate Mechanical Confinement Inhibits Multipolar Cell Division via Pole-Cortex Interaction. *Phys. Rev. X* **13**, 011036 (2023).

178. Silkworth, W. T., Nardi, I. K., Scholl, L. M. & Cimini, D. Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells. *PloS One* **4**, e6564 (2009).
179. Fukasawa, K. p53, Cyclin-dependent Kinase and Abnormal Amplification of Centrosomes. *Biochim. Biophys. Acta* **1786**, 15–23 (2008).
180. Fukasawa, K., Choi, T., Kuriyama, R., Rulong, S. & Woude, G. F. V. Abnormal Centrosome Amplification in the Absence of p53. *Science* **271**, 1744–1747 (1996).
181. Contadini, C. *et al.* p53 mitotic centrosome localization preserves centrosome integrity and works as sensor for the mitotic surveillance pathway. *Cell Death Dis.* **10**, 1–16 (2019).
182. Thompson, S. L. & Compton, D. A. Chromosome missegregation in human cells arises through specific types of kinetochore–microtubule attachment errors. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17974–17978 (2011).
183. Huang, Y. *et al.* Lagging chromosomes entrapped in micronuclei are not ‘lost’ by cells. *Cell Res.* **22**, 932–935 (2012).
184. el-Alfy, M. & Leblond, C. P. Long duration of mitosis and consequences for the cell cycle concept, as seen in the isthmal cells of the mouse pyloric antrum. II. Duration of mitotic phases and cycle stages, and their relation to one another. *Cell Tissue Kinet.* **20**, 215–226 (1987).
185. Baslan, T. *et al.* Ordered and deterministic cancer genome evolution after p53 loss. *Nature* **608**, 795–802 (2022).
186. Collins, R. L. *et al.* A structural variation reference for medical and population genetics. *Nature* **581**, 444–451 (2020).
187. Davoli, T., Denchi, E. L. & de Lange, T. Persistent Telomere Damage Induces Bypass of Mitosis and Tetraploidy. *Cell* **141**, 81–93 (2010).
188. Foulkes, W. D., Flanders, T. Y., Pollock, P. M. & Hayward, N. K. The CDKN2A (p16) gene and human cancer. *Mol. Med. Camb. Mass* **3**, 5–20 (1997).
189. Liggett, W. H. & Sidransky, D. Role of the p16 tumor suppressor gene in cancer. *J. Clin. Oncol.* **16**, 1197–1206 (1998).

References

190. Stefanoudakis, D. *et al.* Significance of TP53, CDKN2A, SMAD4 and KRAS in Pancreatic Cancer. *Curr. Issues Mol. Biol.* **46**, 2827–2844 (2024).
191. Grard, M. *et al.* Homozygous Co-Deletion of Type I Interferons and CDKN2A Genes in Thoracic Cancers: Potential Consequences for Therapy. *Front. Oncol.* **11**, 695770 (2021).
192. Peng, Y. *et al.* Co-occurrence of CDKN2A/B and IFN-I homozygous deletions correlates with an immunosuppressive phenotype and poor prognosis in lung adenocarcinoma. *Mol. Oncol.* **16**, 1746–1760 (2022).
193. Baudis, M. Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data. *BMC Cancer* **7**, 226 (2007).
194. James, M. A., Vikis, H. G., Tate, E., Rymaszewski, A. L. & You, M. CRR9/CLPTM1L Regulates Cell Survival Signaling and is Required for Ras Transformation and Lung Tumorigenesis. *Cancer Res.* **74**, 1116–1127 (2014).
195. Jia, J. *et al.* CLPTM1L promotes growth and enhances aneuploidy in pancreatic cancer cells. *Cancer Res.* **74**, 2785–2795 (2014).
196. Rafnar, T. *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat. Genet.* **41**, 221–227 (2009).
197. James, M. A. *et al.* Functional Characterization of CLPTM1L as a Lung Cancer Risk Candidate Gene in the 5p15.33 Locus. *PLoS ONE* **7**, e36116 (2012).
198. Mardin, B. R. *et al.* A cell-based model system links chromothripsis with hyperploidy. *Mol. Syst. Biol.* **11**, 828 (2015).
199. Zheng, Y.-L., Zhou, X., Loffredo, C. A., Shields, P. G. & Sun, B. Telomere deficiencies on chromosomes 9p, 15p, 15q and Xp: potential biomarkers for breast cancer risk. *Hum. Mol. Genet.* **20**, 378–386 (2011).
200. Perrem, K., Colgin, L. M., Neumann, A. A., Yeager, T. R. & Reddel, R. R. Coexistence of Alternative Lengthening of Telomeres and Telomerase in hTERT-Transfected GM847 Cells. *Mol. Cell. Biol.* **21**, 3862–3875 (2001).
201. M'kacher, R. *et al.* The Transition between Telomerase and ALT Mechanisms in Hodgkin Lymphoma and Its Predictive Value in Clinical Outcomes. *Cancers* **10**, 169 (2018).

202. Zhang, J.-M. & Zou, L. Alternative lengthening of telomeres: from molecular mechanisms to therapeutic outlooks. *Cell Biosci.* **10**, 30 (2020).
203. Turner, K. M. *et al.* Extrachromosomal oncogene amplification drives tumor evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
204. Hull, R. M. *et al.* Transcription-induced formation of extrachromosomal DNA during yeast ageing. *PLOS Biol.* **17**, e3000471 (2019).
205. Yang, L. *et al.* Extrachromosomal circular DNA: biogenesis, structure, functions and diseases. *Signal Transduct. Target. Ther.* **7**, 1–21 (2022).
206. Zuo, S. *et al.* Extrachromosomal Circular DNA (eccDNA): From Chaos to Function. *Front. Cell Dev. Biol.* **9**, (2022).
207. Kang, S. *et al.* TRIM40 is a pathogenic driver of inflammatory bowel disease subverting intestinal barrier integrity. *Nat. Commun.* **14**, 700 (2023).
208. Wu, W., Yang, J., Yu, T., Zou, Z. & Huang, X. The Role and Mechanism of TRIM Proteins in Gastric Cancer. *Cells* **13**, 2107 (2024).
209. Ji, Y. *et al.* Targeting TRIM40 signaling reduces esophagus cancer development: A mechanism involving in protection of oroxylin A. *Int. Immunopharmacol.* **137**, 112362 (2024).
210. Sala, K. *et al.* The ERC1 scaffold protein implicated in cell motility drives the assembly of a liquid phase. *Sci. Rep.* **9**, 13530 (2019).
211. Ma, J., Wang, B., Meng, E. & Meng, X. Case report: identification of ERC1-RET fusion in a patient with pancreatic ductal adenocarcinoma. *Gland Surg.* **10**, 2874–2879 (2021).
212. Feng, G. *et al.* Overexpression of COL5A1 promotes tumor progression and metastasis and correlates with poor survival of patients with clear cell renal cell carcinoma. *Cancer Manag. Res.* **11**, 1263–1274 (2019).
213. Liu, W. *et al.* COL5A1 may contribute the metastasis of lung adenocarcinoma. *Gene* **665**, 57–66 (2018).
214. Tsai, H.-F. *et al.* Type V collagen alpha 1 chain promotes the malignancy of glioblastoma through PPRC1-ESM1 axis activation and extracellular matrix remodeling. *Cell Death Discov.* **7**, 1–12 (2021).

References

215. Gu, S. *et al.* COL5A1 Serves as a Biomarker of Tumor Progression and Poor Prognosis and May Be a Potential Therapeutic Target in Gliomas. *Front. Oncol.* **11**, 752694 (2021).
216. Zhang, J. *et al.* Overexpressed COL5A1 is correlated with tumor progression, paclitaxel resistance, and tumor-infiltrating immune cells in ovarian cancer. *J. Cell. Physiol.* **236**, 6907–6919 (2021).
217. Ren, W. *et al.* Overexpression of collagen type V $\alpha 1$ chain in human breast invasive ductal carcinoma is mediated by TGF- $\beta 1$. *Int. J. Oncol.* **52**, 1694–1704 (2018).
218. Duan, X. *et al.* COL5A1 overexpression correlates with poor prognosis in human cervical cancer. *Int. J. Biol. Markers* **39**, 265–273 (2024).
219. Lane, A. N. & Fan, T. W.-M. Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Res.* **43**, 2466–2485 (2015).
220. Bester, A. C. *et al.* Nucleotide Deficiency Promotes Genomic Instability in Early Stages of Cancer Development. *Cell* **145**, 435–446 (2011).
221. Lalonde, M., Trauner, M., Werner, M. & Hamperl, S. Consequences and Resolution of Transcription-Replication Conflicts. *Life Basel Switz.* **11**, 637 (2021).
222. Mannherz, W. & Agarwal, S. Thymidine nucleotide metabolism controls human telomere length. *Nat. Genet.* **55**, 568–580 (2023).
223. Leonova, K. I. *et al.* p53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E89-98 (2013).
224. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
225. AnnotationDbi. *Bioconductor* <http://bioconductor.org/packages/AnnotationDbi/>.
226. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
227. Morgan, M. & Ramos, M. BiocManager: Access the Bioconductor Project Package Repository. (2024).

228. Access the Bioconductor Project Package Repository.
<https://bioconductor.github.io/BiocManager/>.
229. BiocParallel. *Bioconductor* <http://bioconductor.org/packages/BiocParallel/>.
230. Bioconductor/BiocParallel. Bioconductor (2025).
231. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
232. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).
233. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
234. r-lib/devtools. R infrastructure (2025).
235. Wickham, H., Hester, J., Chang, W., Bryan, J. & RStudio. devtools: Tools to Make Developing R Packages Easier. (2022).
236. Yu, G., Wang, L.-G., Yan, G.-R. & He, Q.-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608–609 (2015).
237. Chang, W. wch/extrafont. (2025).
238. HMMcopy. *Bioconductor* <http://bioconductor.org/packages/HMMcopy/>.
239. Gao, F. *et al.* HTSanalyzeR2: an ultra fast R/Bioconductor package for high-throughput screens with interactive report. in (2017).
240. GenomeInfoDb. *Bioconductor* <http://bioconductor.org/packages/GenomeInfoDb/>.
241. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
242. GenomicAlignments. *Bioconductor*
<http://bioconductor.org/packages/GenomicAlignments/>.
243. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
244. GenomicFeatures. *Bioconductor* <http://bioconductor.org/packages/GenomicFeatures/>.

References

245. GenomicRanges. *Bioconductor* <http://bioconductor.org/packages/GenomicRanges/>.
246. Campitelli, E., Brand, T. van den & olivroy. ggnewscale: Multiple Fill and Color Scales in ggplot2. Zenodo <https://doi.org/10.5281/zenodo.12785523> (2024).
247. Wickham, H. *et al.* Create Elegant Data Visualisations Using the Grammar of Graphics. <https://ggplot2.tidyverse.org/> (2016).
248. ggpubr: Publication Ready Plots - Articles - STHDA. <https://www.sthda.com/english/articles/24-ggpubr-publication-ready-plots/> (2017).
249. ggplot2 Based Publication Ready Plots. <https://rpkgs.datanovia.com/ggpubr/>.
250. KEGGREST. *Bioconductor* <http://bioconductor.org/packages/KEGGREST/>.
251. KEGGREST: Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG) version 1.30.1 from Bioconductor. <https://rdrr.io/bioc/KEGGREST/>.
252. org.Hs.eg.db. *Bioconductor* <http://bioconductor.org/packages/org.Hs.eg.db/>.
253. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
254. Kolde, R. pheatmap: Pretty Heatmaps. (2019).
255. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
256. GitHub - tidyverse/readxl: Read excel files (.xls and .xlsx) into R . <https://github.com/tidyverse/readxl>.
257. Wickham, H. *et al.* readxl: Read Excel Files. (2023).
258. Espejo Valle-Inclán, J. & Cortés-Ciriano, I. ReConPlot: an R package for the visualization and interpretation of genomic rearrangements. *Bioinformatics* **39**, btad719 (2023).
259. team, T. pandas development. pandas-dev/pandas: Pandas. Zenodo <https://doi.org/10.5281/zenodo.13819579> (2024).
260. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
261. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

262. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
263. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
264. Waskom, M. L. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
265. Frank, L. *et al.* ALT-FISH quantifies alternative lengthening of telomeres activity by imaging of single-stranded repeats. *Nucleic Acids Res.* **50**, e61 (2022).
266. Weiskirchen, S., Schröder, S. K., Buhl, E. M. & Weiskirchen, R. A Beginner's Guide to Cell Culture: Practical Advice for Preventing Needless Problems. *Cells* **12**, 682 (2023).
267. DKFZ-ODCF/RNAseqWorkflow. DKFZ - ODCF (2024).
268. DKFZ-ODCF/AlignmentAndQCWorkflows. DKFZ - ODCF (2024).
269. TheRoddyWMS/Roddy-Default-Plugin. The Roddy Workflow Management System (2024).
270. TheRoddyWMS/Roddy-Base-Plugin. The Roddy Workflow Management System (2018).
271. TheRoddyWMS/Roddy. The Roddy Workflow Management System (2024).
272. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
273. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinforma. Oxf. Engl.* **31**, 2032–2034 (2015).
274. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
275. DeLuca, D. S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinforma. Oxf. Engl.* **28**, 1530–1532 (2012).
276. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* **30**, 923–930 (2014).

References

277. Uhrig, S. *et al.* Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res.* **31**, 448–460 (2021).
278. Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757 (2014).
279. Hughes, C. S. *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).
280. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
281. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
282. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
283. Kolberg, L. *et al.* g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–W212 (2023).
284. Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinforma.* **54**, 1.30.1-1.30.33 (2016).
285. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
286. Reisinger, E. *et al.* OTP: An automatized system for managing and processing NGS data. *J. Biotechnol.* **261**, 53–62 (2017).
287. Home. *GitHub* <https://github.com/DKFZ-ODCF/AlignmentAndQCWorkflows/wiki/Home>.
288. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
289. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://doi.org/10.48550/arXiv.1303.3997> (2013).
290. broadinstitute/picard. Broad Institute (2024).

291. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
292. oncoanalyser: Introduction. <https://nf-co.re/oncoanalyser/dev.html>.
293. hmftools/linx/README.md at master · hartwigmedical/hmftools · GitHub. <https://github.com/hartwigmedical/hmftools/blob/master/linx/README.md>.
294. Cameron, D. L. *et al.* GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. *Genome Biol.* **22**, 202 (2021).
295. hmftools/purple/README.md at master · hartwigmedical/hmftools · GitHub. <https://github.com/hartwigmedical/hmftools/blob/master/purple/README.md>.
296. Luebeck, J. *et al.* AmpliconSuite: an end-to-end workflow for analyzing focal amplifications in cancer genomes. 2024.05.06.592768 Preprint at <https://doi.org/10.1101/2024.05.06.592768> (2024).
297. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Comput. Biol.* **12**, e1004873 (2016).
298. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
299. Deshpande, V. *et al.* Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).
300. Seufert, I. *et al.* Enhancing sensitivity and versatility of Tn5-based single cell omics. *Front. Epigenetics Epigenomics* **1**, (2023).
301. shahcompbio/hmmcopy_utils. ShahCompBio (2024).
302. Weber, T., Cosenza, M. R. & Korbel, J. MosaiCatcher v2: a single-cell structural variations detection and analysis reference framework based on Strand-seq. *Bioinforma. Oxf. Engl.* **39**, btad633 (2023).

References

303. Vasimuddin, Md., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (2019). doi:10.1109/IPDPS.2019.00041.
304. Schneider, M. P. *et al.* scAbsolute: measuring single-cell ploidy and replication status. *Genome Biol.* **25**, 62 (2024).