Delineating the Mechanisms through which an Intronic Germline Variant Impacts Acute Lymphoblastic Leukemia Risk

3D genomics revealed an inherited leukemia risk variant to be an enhancer variant associated with *GATA3* upregulation, global chromatin reorganization, and leukemogenesis.

Acute lymphocytic leukemia (ALL) is the most common childhood cancer, with a prevalence of 40 cases per million persons under the age of 15.[1] Although 5-year survival rates have steadily increased over the last few decades, prognosis remains poor for a subset of patients with structural genetic alterations that impact lymphoid development and cytokine receptor and kinase signaling.[2] This high-risk subtype is referred to as Philadelphia chromosome-like (Ph-like) ALL.

Nine genomic loci have been linked to ALL risk by genome-wide association studies: *CDKN2A/2B*, *IKZF1*, *ARID5B*, *CEBPE*, *PIP4K2A-BMI1*, *GATA3*, *TP63*, *LHPP*, and *ELK3*. Variants in the *GATA3* locus have been shown to contribute to Ph-like ALL in particular.[3]

Challenge: Clarifying the Mechanistic Link between a *GATA3* Risk Allele and ALL Pathogenesis

GATA3 variants associated with Ph-like ALL are located in noncoding regions of the genome. Because the function of these regions is poorly understood, pinning down the molecular mechanisms through which these variants contribute to Ph-like ALL is difficult.

To determine how an intronic *GATA3* variant confers ALL risk, researchers led by Feng Yue at Northwestern University explored the 3D genome of mutant and ALL lymphoblastoid cells using Arima technology. Because the variant is located in a noncoding region, they reasoned, it may play a regulatory role in gene transcription by altering local and global 3D genomic interactions.

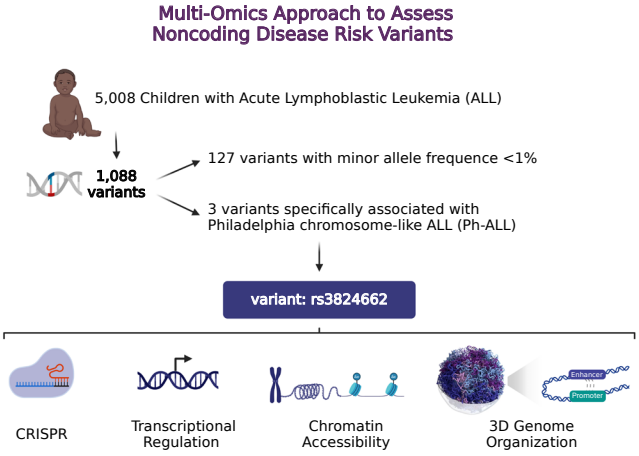


Figure 1. Multi-omics approach used to assess noncoding disease risk variants in Ph-like ALL.[4]

Technology

Arima genome-wide HiC+ enables 3D genome mapping to identify how DNA organization can impact gene regulation and disease processes.

Approach: Leveraging 3D Genomics to Understand the Mechanistic Basis of a *GATA3* Risk Variant

Feng Yue's team previously discovered germline intronic *GATA3* risk variants that increase ALL risk in a GWAS.[4] To comprehensively map *GATA3* risk variants in Ph-like ALL, the team sequenced the *GATA3* locus of 5,008 children with ALL. They identified 1,088 variants, 127 of which had a minor allele frequency of >1%, and 3 of which were specifically associated with Ph-like ALL (versus non-Ph-like ALL). In addition, a *GATA3* variant (rs3824662) with a particularly strong association with Ph-like ALL lined up with a putative enhancer in hematopoietic cells, so they hypothesized that it was a key regulatory variant with potential mechanistic relevance in Ph-like ALL.

To elucidate how variant rs3824662 modifies Ph-like ALL risk, researchers recapitulated the risk A allele in a lymphoblastoid cell line using CRISPR. By comparing risk allele-bearing cells with wild-type cells bearing the reference C allele, the researchers were able to establish the causal effects of the risk genotype. They also validated findings in Ph-like ALL cell lines that bear the A allele.

The researchers first assayed the impact of the A allele on *GATA3* expression. They found that cell lines containing the A allele expressed more *GATA3* than those without an A allele. Subsequent ATAC-seq footprint analysis revealed a proximal NFIC transcription factor binding motif, and a series of reporter, chromatin immunoprecipitation, and gene perturbation assays (targeting NFIC) demonstrated that the upregulation of *GATA3* in A genotypes was a consequence of the relatively higher binding of NFIC to the A allele over the C allele.

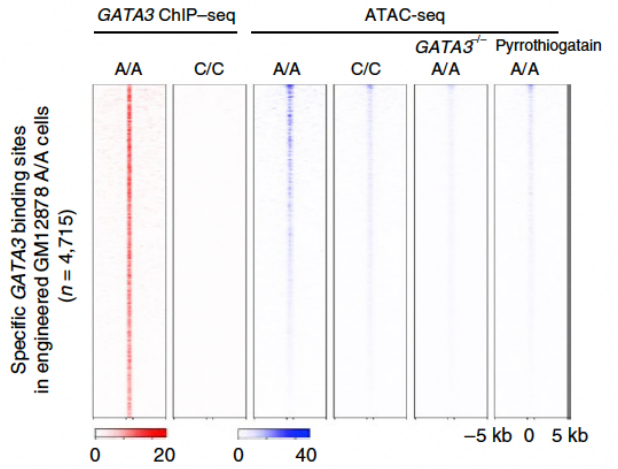


Figure 2. Chromatin accessibility assays highlight widespread binding of *GATA3* to 2,650 chromatin regions that opened after CRISPR knock-in of the Ph-like ALL risk allele.[4]

How does upregulated *GATA3* expression affect global gene expression and chromatin organization? The team approached this question by combining ChIP-seq, ATAC-seq, and differential gene expression analysis. Cells with the A allele displayed a more open chromatin environment, especially around genes that are upregulated in Ph-like ALL. These data hinted at the possibility that *GATA3* is a pioneer factor that opens chromatin around Ph-like ALL loci, driving their expression.

The putative role of *GATA3* as a pioneer factor led the team to investigate how upregulated *GATA3* alters the overall 3D organization of the genome in lymphoblastoid cells. For this task, Feng Yue's team turned to Arima Hi-C. The researchers hypothesized that the extensive gene expression changes observed in A allele lymphoblastoid cells were consequences of loci being reorganized in the two compartments of the genome: the active compartment A and the inactive compartment B. Hi-C showed that replacing the reference C allele with the risk A allele caused 4% of the genome to switch from compartments B to A. Among these changes was the formation of a new chromatin loop between a distal *GATA3*-binding super-enhancer and the CRLF2 locus, which mediates leukemogenesis by constitutively activating JAK-STAT signaling.

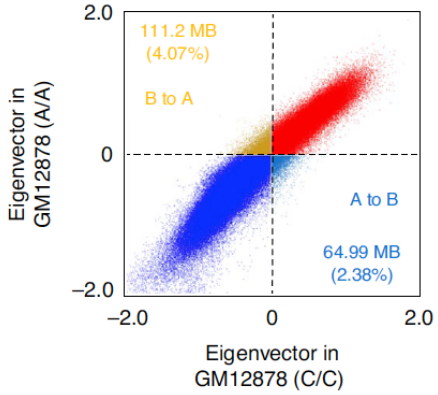
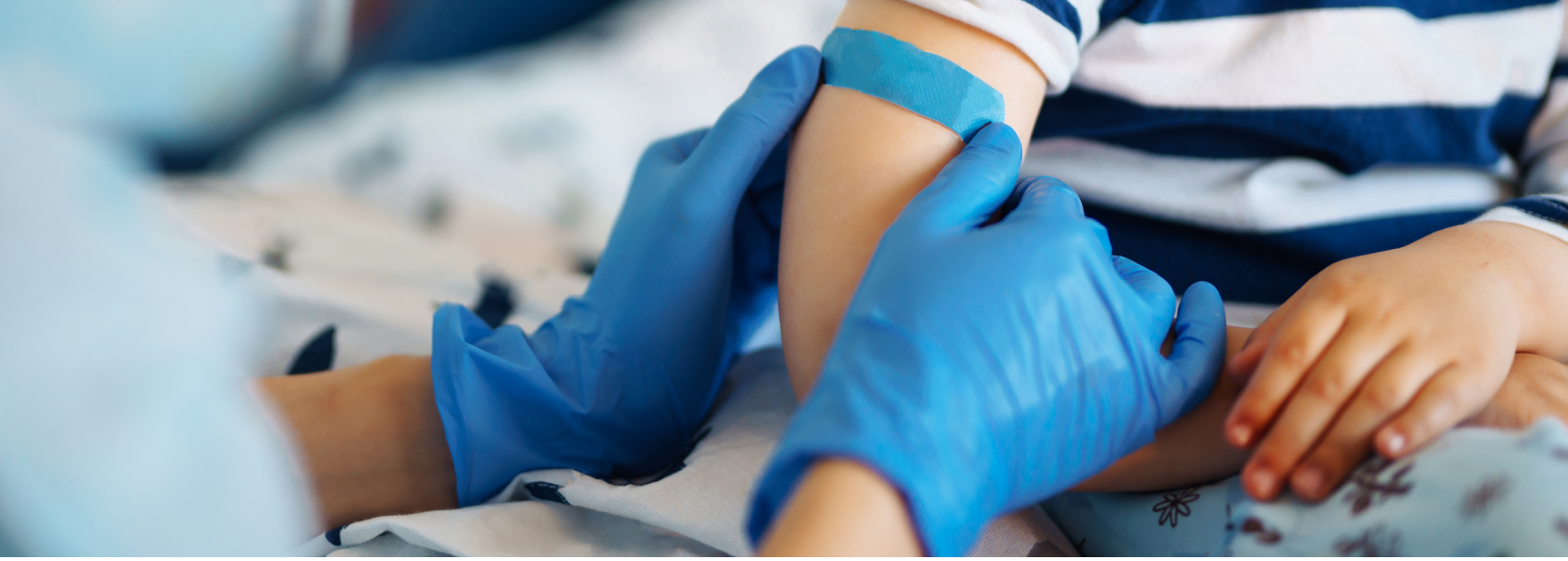


Figure 3. Hi-C revealed substantial compartment switching in the 3D genome of lymphoblastoid cells bearing the Ph-like ALL risk allele.[4]

Finally, the researchers utilized pharmacological and genetic methods to mechanistically tie *GATA3* upregulation to the CRLF2-JAK2 pathway in lymphoblastoid cells. This pathway increased the growth of A allele lymphoblastoid cells in vitro and enhanced their migration in an *in vivo* zebrafish migration assay.



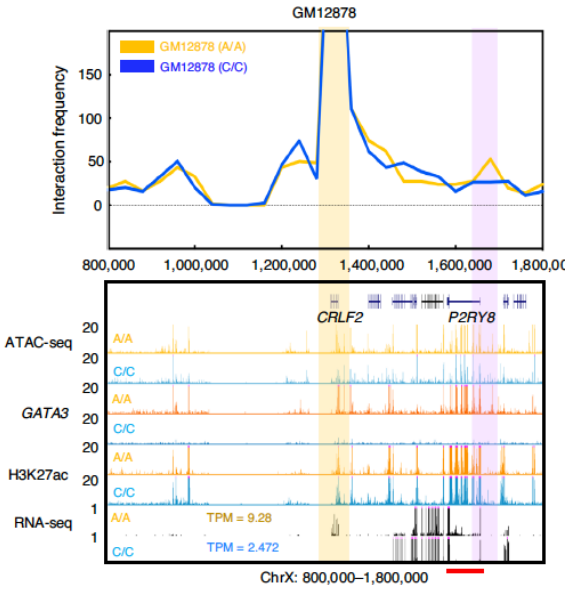


Figure 4. Arima Hi-C defined a new chromatin loop in lymphoblastoid cells bearing the Ph-like ALL risk allele. The loop links the ALL-associated gene CRLF2 (yellow bar) to a *GATA3*-bound super-enhancer (purple bar).[4]

Impact: 3D Genomic Reprogramming Influences Cancer Risk

In this study, the research team identified a noncoding variant strongly associated with Ph-like ALL and uncovered the molecular mechanisms through which it increases Ph-like ALL risk. Although they used a variety of gene perturbation and omics technologies to determine the genomic, epigenomic, and functional effects of the rs3824662 variant, chief among them was 3D genomics using Hi-C, which was instrumental in linking *GATA3* upregulation to CRLF2-JAK signaling.

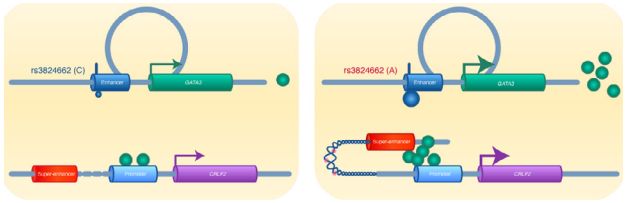


Figure 5. Model illustrating how the rs3824662 variant contributes to Ph-like ALL pathogenesis.[4]

GWAS has identified many cancer variants, but how these variants contribute to cancer pathogenesis remains largely unexplored. 3D genomic techniques such as Hi-C provide a means to delineate the causative mechanisms that link noncoding risk variants to disease pathogenesis, which comprise the vast majority of variants discovered by GWAS. Clarifying these mechanisms may offer clues that could be leveraged for therapeutic and diagnostic strategies for a wide variety of diseases.

References

1. National Cancer Institute. (2022). SEER Cancer Stat Facts: Acute Lymphocytic Leukemia. https://seer.cancer.gov/statfacts/html/alyl.html
2. Den Boer, M. L., et al., (2009). A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: A genome-wide classification study. *The Lancet Oncology*, 10(2), 125–134.
3. Perez-Andreu, V., et al., (2013). Inherited *GATA3* variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nature Genetics*, 45(12), 1494–1498.
4. Yang, H., et al., (2022). Noncoding genetic variation in *GATA3* increases acute lymphoblastic leukemia risk through local and global changes in chromatin conformation. *Nature Genetics*, 54(2), 170–179.