Abstract

Etiologies of pediatric cancers including childhood AML/ALL and brain tumors have been substantially explored. The findings have helped us illustrate a landscape of initiation and progression of tumor genesis, diagnosis/prognosis, and targeted treatments. Owing to the versatility of Next gen sequencing, unprecedented amount of data in cancer patients have been generated in order to characterize the complexity of genetic regulatory mechanisms. Many investigators have started to address cancer-related questions using gene expression profiles, genetic variations, epigenetic profiles, and functional genetic regulatory elements at chromatin level. In this article, we systematically review how various genomic elements, such as gene fusion, mRNA and long non-coding RNA, may contribute to the development of pediatric cancers.

Conclusion: Full understanding of direct/indirect causative and consequent effects of aberrant patterns detected via large-scale platforms in personalized medicine still remain elusive as pieces of puzzles. A better understanding of the role of these different genetic or genomic components may pave the way for identification of novel therapeutic targets.

Keywords: Epigenetics; Mutation; Pathobiology; Pediatric Cancer

Abbreviations:

- AML: Acute Myeloid Leukemia
- AKML: Acute Megakaryoblastic Leukemia
- ALL: Acute Lymphoblastic Leukemia
- CNVs: Copy Number Variations
- CoBRA: Combined Bisulfate Restriction Analysis
- DIPG: Diffuse Intrinsic Pontine Glioma
- EWAS: Epigenome-Wide Association Studies
- GATK: Genome-Analysis Toolkit
- LncRNAs: Long Non-Coding RNAs
- LUNAR1: Leukemia-Induced Non-Coding Activator RNA-1
- Ph-like ALL: Philadelphia Chromosome-Like Acute
- SNVs: Single Nucleotide Variations
- SNPs: Single Nucleotide Polymorphisms
- T-ALL: T cell Acute Lymphoblastic Leukemia
- TSS: Transcriptional Start Site
- 5mC: 5-methylcytosine
- 5hmC: 5-hydroxymethylcytosine
- 5-hydroxymethylcytosine

Background

Clinical heterogeneity has been thought to be a norm rather than an exception in many diseases—which indicates how different deregulated genetic pathways may lead to the same disease. Although many of previous studies[1-19] on the disease spectrum in pediatric AML/ALL and brain tumors have considerably focused on characterization of disease progression and targeted therapy, little has been known how abnormal patterns identified by super machinery Seq-based platforms explicitly work at cellular, molecular, organismal and functional level in...
personalized medicine as the causal and consequent effects. A better understanding of each part of genetic or genomic susceptibility and their ensemble contributors such as gene fusions and disease specific epigenetic modification will enable to more precisely uncover the complexity of pediatric tumorigenesis and target novel treatments. Gene fusions results from the joining of parts of two previously separate genes. It commonly occurs when two different chromosomes break and parts fuse with one another to create a gene that is a hybrid and contains sequences from genes on the two different chromosomes [1,2,3,4,5]. The hybrid gene encodes a fusion protein that contains amino acid sequences derived from the parts of the two genes that were fused. Gene fusions resulting from chromosomal translocations are particularly common in cancers, especially hematological disorders and childhood sarcomas [1,3,6,7,8,9,10,11,12,13].

And expression of genes can be turned on or off or modified by factors other than an individual’s DNA sequence. Although, theoretically, all cells in the same individual contain the same DNA sequences with the exception of mosaic phenomena, the pattern of expression of genes often vary by tissue because of epigenetic regulation [14].

The most common modification that influences expression of genes is DNA methylation. Genes with highly methylated promoters typically are not well expressed. Changes from normal patterns of DNA methylation of specific genes can also cause alterations in gene expression that are associated with disease. Recently it is also shown that DNA and RNA can be modified by methylation at 6' position N6-adenine methylation (N6-methyladenine; 6mA), which may harbor novel roles in gene regulation [5,18]. In addition to modification of nucleic acids, non-coding RNA also serves as another source of epigenetic mechanisms. Emerging evidence has suggested that non-coding RNA such as mRNA and long non-coding RNA (lncRNA) may play a key role in pediatric cancer [10,15,16,17,18,20].

In this article, we systematically review the significant interplay between genomic elements of interest and pediatric cancers that have not been thoroughly investigated thus far, such as gene fusions, epigenetic modifications including small and long non-coding RNAs.

**Architecture of the Genome**

**Gene Fusion**

Identification of the hybrid gene created by gene fusions as cause of a specific type of cancer can lead to the development of effective drugs. An example is the gene created by translocation between chromosomes 9 and 22 to create the “Philadelphia chromosome” characteristic of chronic myelogenous leukemia. The fusion gene (BCR/ABL) encodes an aberrant tyrosine kinase that causes leukemia. It has a catalytic site that differs from normal enzyme. Specific inhibitors of the aberrant tyrosine kinase have been developed that do not inhibit the normal kinase and are effective personalized treatment for the disease [7]. The BCR/ABL fusion ontogeny has also been found in a small fraction of children with B-cell precursor acute lymphoblastic leukemia that has a poor prognosis [21].

The role of fusion genes in disease may be greatly underestimated because most gene expression profile studies have focused on normal versus diseased tissues. Searches for chromosomal translocations and indels using cytogenetic or special microarrays have not always been coupled with studies of gene expression. Recently RNA-seq has been adopted by investigators for transcriptome studies followed by microarray analyses to search for chromosomal abnormalities. A study performed by Atak et al. [22] identified SNVs underlying the over-expression of key driver genes of T cell acute lymphoblastic leukemia (T-ALL) using RNA-seq data. Masettiet al. [10] performed massively parallel sequencing of the transcript me of seven cases of cytogenetically normal pediatric AML and identified a novel recurrent CBFA2T3-GLIS2 fusion, which if present in leukemic cells of a patient predicted a poor outcome. They also identified a new fusion transcript, DHH-RHEBL1, in several CBFA2T3-GLIS2-positive patients. Through the screening of a validation cohort of 55 pediatric AML patients, the authors identified a DHH-RHEBL1 fusion in 8 out of 20(40%) CBFA2T3-GLIS2- rearranged patients. The8 DHH-RHEBL1-positive patients had a poorer survival than the 12 patients whose leukemic cells only harbored the CBFA2T3-GLIS2 rearrangement.

**Epigenetic Modification of the Genome**

**DNA methylation**

Studies of DNA methylation are often coupled with gene expression studies and genetic variation studies, as DNA methylation can regulate gene expression [23]. And SNPs also modify DNA methylation [24,25]. Methylation of DNA involves the conversion of cytosine to methyl cytosine (5mC) by adding a methyl group at 5’ position at a site in DNA, and the oxidative intermediates generated during the de methylation processes(hydroxyl methyl, formyl, and carboxyl-cytosine) [26-28]. The cytosine nucleotide to be methylated is located next to a guanine nucleotide, i.e. in a CpG dinucleotide, although recent research has found methylated cytosine in other sequence contexts, such as Cpa [26].

Epigenome-wide association studies (EWAS) hold promise for the detection of new regulatory mechanisms that may be susceptible to modification by environmental and lifestyle factors affecting disease [27]. Because of the rapid advances in sequencing technology, large numbers of methylated CpG sites can be identified across the entire genome. Studies of cancer frequently compare methylation sites in DNA from cancer tissues with sites in adjacent histologically tumor-free (i.e. normal) tissue. Both cancer
and normal tissues are typically available because of surgical resection of the cancer surrounded by normal tissue margins. DNA methylation profiles have also been used as molecular tools to subgroup cancer patients for personalized treatment [28-30]. In addition, utilizing cell-free DNA methylation as diagnostic tools and markers for treatment efficacy, especially for early detection of cancer, are being rapidly developed [31-33]. However, not many studies have focused on childhood leukemia and brain tumors.

It is important to note that current EWAS studies are limited by methods they used to measure DNA methylation (bead chip vs. sequencing-based methods, comparisons discussed in [34]. Most studies used arrays due to their cheaper cost, ease of use, consistency among large number of samples, and availability of well-established analytical tools. Even if Illumines bead arrays were utilized, the content of probes included (Golden gate vs. 27K vs. 450K) limited the number of CpG sites discovered. Furthermore, most of current studies replied on bisulfate treatment, which studied the sum of 5-methylcytosine (5mC), an oxidized form of 5-methylcytosine, and its base, 5hmC (5-hydroxymethylcytosine), instead of each modification individually.

Non-coding RNA

mRNAs may regulate gene expression by directly interfering transcription, splicing, stability or translation of mRNA [35]. However, lncRNA is likely to regulate gene expression through modulating his tone modification [36], nucleosome positioning [37] and chromatin looping (enhancer RNA) [38-41]. As a result, they can target tumor-related signaling pathways, regulate inflammation and senescence-related genes [20], and promote proliferation and tumor growth by modulating cell cycle and inhibiting apoptosis [19, 42-43]. The roles of lncRNA in other aspects of cancer development have been extensively discussed [44]. Moreover, the use of mRNA and lncRNAas diagnostic and prognostic markers for pediatric tumors has been suggested [45,46].

In addition to pediatric brain tumors, emergent evidence also suggests the roles of mRNAs and long non-coding RNAs in T-ALL and AML. Trimarchiand colleagues (2014) [47]. Performed a relatively comprehensive characterization of long non-coding RNAs and proposed them as key multi-players leading to tumor genesis such as cellular homeostasis or transformation. In the study, a representative IncRNA referred to as LUNAR1 (LeUkemia-induced Non-coding Activator RNA-1) has been thoroughly explored for its role in regulating neighboring genes such as IGF1R gene to promote Igf1 signaling and the growth of T-ALL by altering 3D chromatin structure and enhancer-specific features. Another study by Hirano et al. (2015) [48] has highlighted that IncRNA CCDC26 could be an effective therapeutic target to control myeloid leukemia cell growth via regulation of KIT expression, which is related to cell proliferation or survival. Furthermore, Emmrich and colleagues [49] found that an interconnected functionality of lncRNAs MONC and MIR100HG plays a critical role as regulators of hematopoiesis and oncogenes, contributing to the maintenance of leukemic growth and therefore leading to the progressive Acute Megakaryoblastic Leukemia (AMKL). However, studies on mRNA and IncRNA at a genome-wide scale are still scarce and more comprehensive profiling is needed to further understand their roles in functional significance of cancer development, to evaluate their usage as diagnostic and prognostic markers, to develop more sophisticated and precise personalized treatments, and to determine subtype-specific features in tumor genesis [50,51]. In the following section, we will further discuss the role of DNA methylation in pediatric leukemia and brain tumor.

Disease-specific DNA methylation

The role of DNA methylation in cancers as well as diseases associated with imprinting has been extensively studied. With the emergence of high-throughput technologies, ranging from microarrays to next-gene sequencing, genome-wide scans to search for disease-related DNA methylation markers is now possible [52-58]. The interplay among DNA methylation sites, genetic variation, protein binding sites, and gene expression is a very active field of investigation.

Childhood leukemia

Nordlund et al identified genome-wide signatures associated with pediatric acute lymphoblastic leukemia (ALL) in 764 children at diagnosis and 27 children at relapse using the Infinium Human Methylation 450k Bead Chips [56]. In addition to 9,406 hyper methylated CpG sites in ALL cells independent of cytogenetic background, each cytogenetic subtype displayed a unique set of hyper- and hypo methylated CpG sites. Further, these investigators identified DNA methylation signatures in cells at relapse that differed from those at the initial diagnosis from patients. Changes in DNA methylation signatures can predict disease relapse. Subtype-specific methylation patterns in promoters and enhancers were strongly correlated with gene expression. Hypo- and hyper-methylated sites are generally over-and under-expressed, respectively. Wong et al [58] studied DNA methylation profiles in archived bone marrow smears from 19 children with ET维尔-RUNX1 (fusion gene that promotes survival of early lineage B-cells) positive pediatric pre-B cell ALL. Smears had been collected at diagnosis and remission and were used to derive a disease-specific DNA methylation profile. Gene signatures were confirmed in an independent cohort of 85 patients by quantitative analysis of DNA methylation and identification of differentially methylated CpG sites using SEQUENOM Epityper. Methylation patterns at 15 loci were sufficient to discriminate leukemic from disease-free samples, regardless of the cytogenetic subtype of pre-B ALL. Aberrant methylation of TCF3, EGR4, BTG4 and hypomethylation of POU2AF1, TCF3, and RAG1 were found.
in leukemic cells of ETV6-RUNX1 positive pediatric pre-B cell ALL cases. These genes are associated with B cell development. Chatterton et al [59] studied the DNA methylation and gene expression in 69 pediatric B-cell ALL patients (26 ETV6-RUNX1, 15 hyper diploids, and 28 “B-cell other”) and 48 non-leukemic control samples. They identified 795 CpG sites with similarly deregulated DNA methylation levels common to three B-cell ALL subtypes, correlated with expression changes in 370 genes. In addition, there were also subtype-specific DNA methylation alterations with concurrent gene expression changes, which would help to increase understanding of the pathogenesis of each disease subtype. They can potential guide refinement of current treatment regimes tailored to each individual and can be used as robust biomarkers of disease progression states.

Figueroa et al. [60] studied the interaction among epigenetic markers, gene expression, and copy number variations in 137 B-lineage and 30 T-lineage childhoods lymphoblastic leukemia’s using microarray-based technology. Consistent with other studies, this group found that different genetic subtypes of ALL were characterized by distinct DNA methylation signatures that exhibit significant correlation with gene expression profiles. There were epigenetic signatures common to all cases, with correlation to gene expression in 65% of these genes, suggesting that a core set of epigenetically deregulated genes is central to the initiation or maintenance of lymphoid transformation. Finally, genes with aberrant methylation were also targeted by recurring DNA copy number alterations in ALL, which suggests that expression of genes in CNVs are changed far more frequently than suggested by structural genomic analyses alone.

**Childhood brain tumors**

Although brain tumors are not as common as blood cancers in children, pediatric-onset brain tumors are characterized with a highest cancer-related mortality rate in children. The high mortality rate may be at least partly attributable to a limited number of clinical trials due to the scarcity of cases [61,62]. Furthermore, heterogeneous genetic mutations associated with some pediatric-onset brain tumors may also partially explain why some single therapeutic agent cannot improve the clinical condition for a majority of cases. Omics-data facilitated by high-throughput sequencing have produced expansive evidence for novel epigenetic mechanisms underlying brain cancers, such as histone mutations, hijacking of enhancer elements, and novel oncogenic gene fusions - to name a few of them. For example, structural variants near the growth factor independent 1 family proto-oncogenes, GF11 and GF11B, may be active enhancer elements, including super-enhancers to cause medulloblastoma [59]. These findings exemplify how heterogenous the genetic factors associated with the susceptibility to pediatric-onset brain tumors could be.

The role of DNA methylation in childhood neuroblastoma has been explored by Lau and colleagues using customized Illumine Golden Gate methylation assays for 96 CpG sites within 48 candidate genes in primary neuroblastoma tumors from 131 children [60]. Levels of DNA methylation were validated in a subset of 48 patients using combined bisulfate restriction analysis (CoBRA) and bisulfate sequencing. The authors discovered that hyper methylation of FOLH1, MYOD1 and THBS1 are significant independent predictors of poorer clinical outcomes after adjusting for known prognostic factors. This study highlights the potential use of methylation profiling to identify prognostic markers and detect new therapeutic targets for selected subsets of patients. Studies of DNA methylation are often coupled with gene expression studies and genetic variation studies, as DNA methylation can regulate gene expression [63] and SNPs also modify DNA methylation [64-65]. Hovestadt et al. used Infinium Human Methylation 450k Bead Chips to study medulloblastoma and have successfully identified 48 CpG sites whose methylation can distinguish 4 subgroups of medulloblastoma [34]. Danielsson et al. further extended this study to all pediatric brain tumors and have developed a tool named MethPed that allow researchers and clinicians to classify patients into 9 brain tumor subtypes for further consideration of treatment development [37]. Fontebasso et al. combined whole exome sequencing, gene expression and global DNA methylation profiles to study pediatric midline high-grade astrocytomas (including diffuse intrinsic pontine glioma or DIPG) [64]. Mutations in his tones have been found common to most tumors (37/40), which are correlated with common changes in DNA methylation. Recently, the deregulation of Loss of 5hmC in DIPG has been suggested [65] and genome-wide studies are hence warranted to examine the roles of 5mC and 5hmC in brain tumors.

**Gene Expression Profiles**

Comparing gene expression profiles of normal with diseased tissues has long been a common research method to study the pathophysiology of disease. Originally such studies were primarily conducted using expression microarrays. Use of microarrays is now typically coupled with studies using high throughput technologies such as RNA-seq[70-78]. These techniques have enabled the generation of lists of top putative candidates of differentially expressed genes between different groups in diseases of interest. Studies of gene expression generate information about alternative splicing of RNA transcripts, the role of non-coding genomic elements such as ncRNAs, ncDNAs, and microRNAs and epigenetic changes in the genome [52-54,66-72].

Leukemia’s are the most common pediatric cancers. Several groups have studied gene expression profiles in B-cell and T-cell pediatric acute lymphoblastic leukemia at diagnosis and relapse. And researchers have conducted
integrative approaches between gene expression and epigenetic modifications of the genome [24,70,73-76]. Busche et al. [24] focused on pre-B ALL carrying the ETV6-RUNX1 fusion gene created by translocation between chromosomes 12 and 21. This study aims at profiling a meta-data between DNA methylation and transcript me. They demonstrated that genome-wide DNA methylation profiling could separate pre-B ALL into major subtypes with different responses to therapy and frequency of relapse. The quantification of methylation levels in CpG-islands around the transcriptional start site was statistically compared in Spearman correlation measurement with expression levels, resulting from the varying spectrum of correlations for each patient sample. Interestingly, they compared methylation levels with RNA-seq gene expression data in 17 patient samples and observed inverse correlation between methylation at CpG-sites located within the TSS1500 and its corresponding gene expression levels. They found the highest inverse correlations for CpG-sites located within exon1, and the next highest inverse correlations, in turn, at TSS (CpG sites within 1kb from transcriptional start site), and 5’-UTR, whereas it shows slightly significant positive correlations at 3’-UTR. Correlation analysis irrespective of subtype-specific molecular features demonstrated differential methylation of previously identified sub-type classifier for t (12; 21) such as BMP4, CELSR1, DSC3, and PON2 that are significantly correlated with transcript expression in ALL. Thus, they demonstrated that a combined strategy between methylene and transcript me analyses can identify potential drivers of leukemogenesis. Meyer et al. [77] generated transcript me profiles of matched diagnosis and relapse bone marrow specimens from ten individuals with pediatric B-ALL using RNA-seq and identified Single Nucleotide Variations (SNVs) and insertions/deletions (indels) by the means of Genome-analysis Toolkit (GATK) and fusions by utilizing in-house pipeline, respectively. They identified 20 newly acquired non-synonymous mutations not present at initial diagnosis. At relapse two of the 10 patients had acquired new mutations in the 5’-nucleotidase NT5C2. Full-axon sequencing of NT5C2 was completed in 61 further relapse specimens, 5 of which had a new mutation in NT5C2. Collectively, the study showed that all individuals who harbored NT5C2 mutations relapsed early, indicating that mutations in NT5C2 are associated with the outgrowth of drug-resistant clones in ALL and a poor prognosis. Identification of epigenetic and genomic alterations that change expression profiles of genes in cancer cells from patients who have responded poorly to current therapies help identify targets for development of new drugs. More recently, Roberts et al. [78] explored Philadelphia chromosome-like acute lymphoblastic leukemia (Ph-like ALL) on the basis of Next gen transcript me, whole-genome, exome sequencing, microarray profiling, and functional and cytogenetic assays. In the study, they pointed out that kinase-activating alterations in more than 90% of patients with Ph-like ALL were observed with rearrangements and sequence mutations and the authors suggest that adding tyrosine inhibitors to current therapy will be a better approach to increase the survival rate of this type of patients.

Genetic Regulatory Networks

Information on co-regulators and co-expression of genes in normal and disease states has increased interest in genetic regulatory networks. The generation of genetic variation/gene expression/DNA methylation profiles by either array or sequencing platforms provides the opportunity to define connectivity maps. This permits assignment of genetic variants, differentially expressed or methylated genes associated with pediatric disorders to specific biochemical pathways and regulatory networks [79-85]. These technologies have been applied in cancer and other diseases and have uncovered genetic regulatory mechanisms of action for targeting disease-specific genes. The complex underlying mechanisms and phenotypic outcomes for diseases such as cancer are derived through multi-layers of biological processes involved in development, cell-cycle, and epigenetic machinery with cis-regulating and trans-acting effects [4,24,26,52,53,55-60,67,68,70,71,73,76,79,80,82,83,86-99]. Foreexample, large-scale sequencing of primary tumors from childhood medulloblastoma patients and their matched normal blood revealed that mutations in enzymes responsible for epigenetic and histone modifications are present in all subtypes of this disease, suggesting that these mutations are prime candidates of driver events [100-105]. This led to identification of genes and pathways that might be associated with poor outcomes and could be targets for new drugs to treat medulloblastomas with specific genetic mutations. Transcriptional profiling of medulloblastoma cohorts identified several distinct molecular subgroups that vary with respect to their demographics, responses to therapy, and clinical outcomes [101,106-109]. Neuroblastoma, like medulloblastomas, is a cancer of nerve cells and has been studied in a similar fashion [110,110,111]. Molecular classification of medulloblastomas and neuroblastoma is possible and similar approaches have started to be applied to other pediatric diseases. Knowledge of genomic changes can be used to identify the biochemical pathways and regulatory networks that are affected and to identify potential targets for drug therapy [98,112]. Molecular classification of a patient’s disease should improve choices of therapy at diagnosis, given a particular molecular subtype, and lead to development of new therapies for subtypes with a poor prognosis.

Concluding Remarks

The rapid advances of genomics genetics research have shed light on several novel therapeutic targets for pediatric cancers. In this systematic review, we have comprehensively reviewed various genomic elements associated with pediatric cancers. We have focused on ontogeny driver genes, mutations, epigenetic modifications regulating gene expression in protein-
coding genes and non-coding regions, aberrant patterns on gene expression profile data, post-transcriptional splicing procedures and chromosomal translocations of fusion genes. These genetic or genomic elements may serve as indicators to evaluate clinical stages and treatment options and responses. The accurate protocols for detection of diagnosis and prognosis biomarkers and treatments for specific cancers are still unclear and the well-defined sub-classification of tumor types is generally lacking at present [113-115]. In addition, it still remains unclear how disease-associated variants, gene expression profiles, epigenetic patterns, and other genomic phenomena, influence each other in cells. It is still not clear to define explicitly what roles the abnormal patterns have in order to derive the complex multiple-steps in initiation and progression of tumors [23,94,100,102,103,116].

Above and beyond targeting genes and mutations at individual level, in the state-of-principle, further development of methods in system biological approaches to integrate single-layer analyses is crucial [8,23,79,117-119]. More recently, a few studies have investigated pediatric cancers by integrating methylome and transcriptome analyses, combining analyses of whole genome-wide association studies/exome sequencing and expression profiles, exploring micro RNAs-RNAs interactions, and integrating transcriptome and proteomics studies. In order to decode the entire cascade of tumor genesis that is induced by accumulating effects from multi-layer events, it is not sufficient to study the intricate cancer biology with the regards to its functional roles by quantification at the single level alone.

There currently exists a significant gap to define the relationship between the changes of genomic abnormalities and their impact on tumor growth signaling pathways and responses to external stimuli such as targeted drugs. Hence, it is also very important to precisely account for appropriate experimental settings and validation steps, analytical and methodological pipelines, and well-balanced clinical applications in cohort studies with suitable sample sizes and proper selection of clinical informative measurements. Thus, above of all, we need efficient and systematic strategies to characterize deregulated physiological functions caused by somatic mutations, oncogenic fusions, and other abnormalities in genomic elements at sub-molecular, cellular, tissue-specific and organismal levels during the initiation and progression of pediatric tumors. A better understanding of the landscape of various genetic functional changes associated with pediatric oncogenesis will provide an initial key to personalized medicine.

References


