Genes and pathways underpinning Klinefelter syndrome at the single-cell level

Linlin Tian
Nanjing Municipal Center for Disease Control and Prevention

Yan Yu
Nanjing Municipal Center for Disease Control and Prevention

Ziqing Mao
Nanjing Municipal Center for Disease Control and Prevention

Dandan Xu
Nanjing Municipal Center for Disease Control and Prevention

Hongbo Zhang
Nanjing Municipal Center for Disease Control and Prevention

Mengkai Qiao
Nanjing Municipal Center for Disease Control and Prevention

Tong Chen
Nanjing Municipal Center for Disease Control and Prevention

Wen Liu
liuwen0905@126.com
Nanjing Municipal Center for Disease Control and Prevention

Research Article

Keywords: KS, hub genes, pathways

Posted Date: August 29th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3285636/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.
**Version of Record:** A version of this preprint was published at Biochemical Genetics on February 19th, 2024. See the published version at [https://doi.org/10.1007/s10528-024-10689-6](https://doi.org/10.1007/s10528-024-10689-6).
Abstract

Klinefelter syndrome (KS) is the most frequent genetic anomaly in infertile males. Despite this, the molecular mechanisms involved in KS are poorly defined. Based on bulk transcriptome and single-cell RNA sequencing datasets with peripheral blood monocyte (PBMC) sample(s) from healthy and KS men, this study was designed to address critical genes and pathways correlated with the occurrence of KS. Through a comparison between control and KS samples, we obtained 5 hub genes, including two upregulated genes (XG and ITLN1) and three downregulated genes (DEFA4, BPI and MPO). Without exception, these five genes yielded an excellent discriminatory capacity for KS with an area under the receiver-operator-characteristic curve over 0.75. We also assessed between-group differences of immune cell infiltration using ssGSEA. Infiltrated degree of some immune cells such as CD56\textsuperscript{bright} NK cell was found to be positively associated with the expression of ITLN1 and XG. Through Kyoto Encyclopedia of Genes and Genomes enrichment, we identified PI3K/AKT pathway and neuroactive ligand-receptor interaction as upregulated pathways for KS. Gene set enrichment analysis together with gene set variation analysis confirmed upregulation of G2M checkpoint, mitotic spindle, and heme metabolism for KS. Furthermore, scRNA-seq data analysis was conducted to detect intercellular communication between different immune cell types, and a strong correlation was detected for macrophages, dendritic cells or NK cells with the other cell types. Collectively, we provided hub genes, pathways, immune cell infiltration degree, and cell-cell communication interactions for KS, warranting novel insights into the mechanisms of KS.

Introduction

First described by Klinefelter H. in 1942 (Klinefelter 1942), Klinefelter syndrome (KS) is the most frequent type of sex chromosomal abnormality in human males. KS affects 1 in 650 male births in the normal population (Kanakis and Nieschlag 2018) and accounts for as high as 12.62% of men with nonobstructive azoospermia (NOA) (Haihong et al. 2014). Approximately 80–90% of KS men are non-mosaic with a 47,XXY karyotype, while the remaining 10–20% of KS men are mosaic or accompanied by higher grade of aneuploidy (Bonomi et al. 2017). The predominant presentations of KS consist of small-volume testis, hypergonadotropic hypogonadism, as well as declined semen quality (Bonomi, Rochira, Pasquali, Balercia, Jannini and Ferlin 2017, Groth et al. 2013, Lanfranco et al. 2004). Advanced maternal age is the only evidence-based risk factor for KS (Tüttelmann and Gromoll 2010). With the advancing technology in reproductive medicine (for example, testicular sperm extraction followed by intracytoplasmic sperm injection), it has become possible for approximately half of men with KS to obtain biological offspring(s) (Bhasin and Oates 2020, Zganjar et al. 2020). In doing so, a significant number of KS patients are thus forced to select assisted reproductive technology using donor spermatozoa or other limited options. Thus, understanding regarding the prevention or treatment of KS is in urgent need.

Undoubtedly, a clearer understanding of the etiology of KS will provide a more optimized diagnostic and therapeutic strategy for this intractable disorder. In up to half of KS patients, the onset of this disorder is
attributed to non-disjunction at the time-point of meiosis I during paternal spermatogenesis (Hassold et al. 1991, Thomas et al. 2000). In doing so, a significant number of parental spermatozoa of KS patients exhibit an abnormal XY karyotype (Eskenazi et al. 2002, Lowe et al. 2001). Further, one of the leading causes of paternal-origin KS is impaired recombination between paternal sex chromosomes within the pseudo-autosomal region during meiosis I (Hassold, Sherman, Pettay, Page and Jacobs 1991, Thomas, Collins, Hassold and Jacobs 2000). One recent research showed that paternal USP26 mutation increases KS incidence of offsprings in both mice and humans (Liu et al. 2021). KIF2C plays an critical role in the differentiation of spermatozoa in KS men (He et al. 2022). Immune activation is found both in the testis and in the peripheral blood of KS men (Zhao et al. 2020), implying immune-related factors essential for the onset of KS. Also, like the immune anomalies, metabolic dysfunctions are involved in the pathogenesis of KS (Liu, Tang, Zheng, Xu, Guo, et al. 2019). Despite this enormous progress, a complete understanding of molecular mechanism that underpin KS-related infertility remains missing.

In view of above considerations, the present study integrated one single-cell RNA-seq (scRNA-seq) and two bulk microarray transcriptome datasets. Specifically, after batch effect removal, two bulk transcriptomic datasets were merged to identify the hub genes between control and KS samples. Then, we applied receiver operating characteristics (ROC) curve for the evaluation of the diagnosis performances of hub genes for distinguishing control and KS samples. We also investigated critical terms and pathways of KS using Gene set enrichment analysis (GSEA), gene set variation analysis (GSVA), Gene Ontology (GO), and Kyoto Encyclopedia of Genes (KEGG) analysis. Besides, we applied single sample gene set enrichment analysis (ssGSEA) for detecting infiltrated immune cell abundances. Moreover, one scRNA-seq transcriptomic dataset containing peripheral blood monocyte (PBMC) sample from KS patient was performed for in-depth analysis of cell-cell communication among different immune cell types. Overall, our findings provide new insights into putative biomarkers and therapeutics for KS.

**Materials and Methods**

**Data acquisition**

We obtained GSE42331 (Michael et al.), GSE47584 (Liang et al. 2015) and GSE136353 (Liu, Tang, Zheng, Xu and Dai 2019) datasets from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). Data inclusion criteria were defined as follows: (a) the organism was homo sapiens; (b) peripheral blood Monocyte (PBMCs) samples incorporated control and KS samples; (c) data for control and KS samples were intact; (d) control and KS samples were explicitly separated via principal component analysis (PCA). One dataset was included in the current study when all the above requirements were met. GSE42331 containing 15 control cases and 35 cases of KS was from GPL6244 platform (Affymetrix Human Gene 1.0 ST Array version). The GSE47584 containing 5 control cases and 5 KS cases was acquired from GPL14550 platform (Agilent-028004 SurePrint G3 Human GE 8x60K Microarray Probe Name Version). GSE136353 containing PBMCs sample from one KS patient was acquired from the GPL24676 Illumina NovaSeq 6000 (Liu, Tang, Zheng, Xu and Dai 2019).
Data pre-processing and differentially expressed genes (DEGs) detection

Following probe transformation, we obtained corresponding gene symbols. When encountered a correspondence between one gene symbol several probes, we regarded the first expression level of this gene as the applied value. Data were subsequently filtrated, log 2 -transformed, and normalized. After merging, we applied ComBat function of R package sva (Leek et al. 2012) to correct batch effect, followed by the identification of sample distribution through principal component analysis (PCA). A comparison of control and KS samples was made through R package limma to find DEGs (Ritchie et al. 2015), with the cut-off criterion false discovery rate (FDR) below 0.05 and |log2FC| exceeding 0.5.

Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA)

We performed GSEA using clusterProfiler (Yu et al. 2012) and adjusted P value through the Benjamini-Hochberg (BH) method. We then selected FDR < 0.05 as the threshold for determining the hallmark pathway significantly enriched. Besides, we also applied GSVA for the identification of significant hallmark pathways. Immune-related pathways were explored using GSVA (Hnzelmann et al. 2013). The gene set “h.all.v7.1.symbols.gmt” from the Molecular Signature Database was set as a reference.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis

We conducted GO and KEGG analysis through clusterProfiler (Yu, Wang, Han and He 2012) for gene functional annotation, and the “gseGO”and“gseKEGG” function was used to screen related pathways. GO terms consist of biological process (BP), cellular component (CC) and molecular function (MF). FDR < 0.05 was regarded as significantly enriched.

Immune cells infiltration analysis

For the assessment of enriched score of PBMC samples, we implemented ssGSEA through GSVA (Hnzelmann, Castelo and Guinney 2013). Correspondingly, we chose gene expression values of PBMC samples and metagenes of 28 immune cell types. We took normalized enrichment score (NES) as immune-infiltrated level. The correlation of hub genes and the immune cells were also computed.

ScRNA-seq transcriptome analysis

We analyzed the ScRNA-seq transcriptome through Seurat (Stuart et al. 2019). Subsequently, we filtered out cells with nFeature_RNA ratio ≤ 200 or mitochondria ratio ≥ 10%. We then reduced data dimension through UMAP. A total of 6723 PBMC cells from the KS patient were detected to generate different clusters. All cells were clustered based the marker genes of each cell type: NK cell (NKG7), T cell (CD3D and CD3E), B cell (CD79A and CD79B), macrophage (CD14) and dendritic cell (CD33). The
communications of different cell types were further investigated through the R package CellChat (Jin et al. 2021).

Statistical analysis

We conducted data analysis using R software. The ROC curve was obtained through plotting sensitivity as well as 1-specificity for each cut-off. We applied receiver operating characteristic (ROC) curve to evaluate the specificity along with sensitivity of hub genes. The area under curve (AUC) > 0.70 and P value < 0.05 indicated statistical significance.

Results

Data preprocessing and DEGs identification

The expression profile of two bulk datasets was integrated. We subsequently log-normalized the data in each bulk dataset, followed by the correction of batch effect (Figs. 1A, B). We applied PCA for the identification of sample distribution before and after batch effect correction (Figs. 1C, D). It could be detected that sample distribution was more consistent following the elimination of batch effect. Based on the comparison of samples between the control and KS groups, we identified 5 DEGs (2 upregulated and 3 downregulated genes), which were regarded as the hub genes (Fig. 1E).

Validation of KS-related hub genes

ROC curves displayed the diagnostic accuracy of the hub genes to distinguish between the control and KS samples. The AUC for the generated ROC were greater than 0.75 without exception. To be specific, the AUC was 0.786 (95% CI, 0.653–0.918) for DEFA4 (Figs. 2A), 0.803 (95% CI, 0.678–0.926) for BPI (Figs. 2B), 0.816 (95% CI, 0.695–0.937) for MPO (Figs. 2C), 0.756 (95% CI, 0.620–0.891) for ITLN1 (Figs. 2D) and 0.776 (95% CI, 0.653–0.899) for XG (Figs. 2E). Therefore, high diagnostic accuracy was observed for each hub gene to diagnose KS.

Aberrantly regulated hallmark pathways in KS samples

We firstly investigate significantly enriched hallmark pathways for KS through GSEA (Figs. 3A-C). The upregulated hallmark pathways for KS included G2M checkpoint, heme metabolism, interferon alpha response, interferon gamma response, and mitotic spindle. Additionally, we presented the genes specific to the top enriched hallmark pathways (Fig. 3B). Moreover, GSVA was performed, and correlations of hallmark pathway expression levels with samples were visualized (Fig. 4A). The significantly dysregulated hallmark pathways in KS are shown in Fig. 4B. Then, GSEA and GSVA derived pathways were intersected to find 3 upregulated and 11 downregulated pathways (Fig. 5).

GO and KEGG enrichment
For further exploring biological function and related pathway in association with KS, GO analysis was performed, and genes mainly enriched in DNA recombination, cation channel complex, chromosomal region and hormone activity (Fig. 6A, C, E). To depict functionally enriched GO terms, we utilized upset plots to reveal correlations of specific genes with enriched terms (Fig. 6B, D, F). Additionally, KEGG analysis was performed, and genes were markedly enriched in neuroactive ligand-receptor interaction, PI3K/AKT signaling pathway, etc (Fig. 6G-H).

**Immune cell abundance**

To explore immunological changes during KS progression, we utilized the ssGSEA for the evaluation of immune cell infiltrated levels in control and KS samples. The expression correlations of infiltrated immune cells in control or KS samples were described (Fig. 7A). The infiltrated correlations among different immune cells were detected for the control and KS groups, respectively (Fig. 7B). Moreover, we computed the associations between hub genes and the immune cells (Fig. 7B). It should be noted that, ITLN1 and XG were positively associated with CD56<sup>bright</sup> NK cell concurrently.

**Single-cell profiling and cell–cell communication**

We obtained a total of 6,723 single-cell transcriptomes from the KS patient in the scRNA-seq dataset and detected 9 cell clusters (Fig. 8A). The cluster specific genes were compared with the previously reported cell type-specific marker genes to identify the cell type of each cluster. In doing so, we detected five cell types (Fig. 8B), including T cells (Clusters 1, 4, 3, 7), NK cells (Clusters 0, 6), B cells (Cluster 2), macrophage cells (Cluster 5), and dendritic cells (Clusters 8, 9). The presentative cell type-specific marker gene exhibited correspondingly accurate and specific expression distributions (Fig. 8B). We also explored interaction networks and differential ligand-receptor pairs among different immune cells. The global cell interaction network together with each cell type-specific interaction network were visualized (Fig. 8C). It could be found that B cells and T cells displayed weaker interactions with the other cell types.

**Discussion**

This study collectively integrated three transcriptomic datasets with PBMCs samples from KS patients to explore the potential pathogenesis of KS. Based on two bulk transcriptomic datasets, we detected hub genes for KS, including DEFA4, BPI, MPO, ITLN1 and XG. DEFA4, one family member of defensins, is the component of antimicrobial peptide of the innate immune system. DEFA4 is predominantly expressed in human bone marrow and whole blood. DEFA4 upregulation has been found to be associated with multiple immune-related diseases. For instance, elevated expression level of DEFA4 was reported to be correlated with the severity of COVID-19 symptoms (Overmyer et al. 2021). Furthermore, DEFA4 was found to have antiviral effect, which act efficiently against HIV-1 (Human neutrophil alpha-defensin 4 inhibits HIV-1 infection in vitro 2005). BPI is a single-chain cationic protein that belongs to a conserved lipid-transfer protein family. In the male mice, BPI is specifically expressed in the testis and epididymis (Lennartsson et al. 2005). After secretion from the epididymal epithelia, BPI is transferred to membrane
surface of the spermatozoa (Zhou et al. 2014), and is essential for the sperm-oocyte fusion (Li et al. 2013, Yano et al. 2010). MPO serves as peroxidase lysosomal enzyme that produces hypochlorous acid and other highly reactive oxidants (Klebanoff et al. 2012). Previous study showed that extracellular levels of MPO in the seminal plasma were negatively associated with sperm concentration and motility in men (Pullar et al. 2017). Moreover, neutrophil was reported to produce and release oxidative and cytotoxic molecules such as MPO into spermatogenic microenvironment, which in turn lead to testis tissue damage. Endosulfan exposure could damage murine testis through upregulating extracellular MPO levels and inhibiting endogenous antioxidant enzymes (A et al. 2021, Nna et al. 2019, Tesi et al. 2022). ITLN1, one adipocytokine secreted in the seminal vesicles, can effectively eliminate inflammatory response and oxidative stress. ITLN1 plays important roles in energy homeostasis, glucose metabolism, and cardiovascular protection (Takuya et al. 2017). One recent study proposed that seminal ITLN1 levels were decreased in patients with varicocele, leukocytospermia or smoking habit than in fertile men (Ismail et al. 2017).

In terms of functional enrichment, enriched GO terms contained tumor necrosis factor processes (Fig. 6A). TNF-α acts as a multifunctional cytokine in mammalian testes and critical for spermatogenesis and male fertility (Birt et al. 2013). In accordance with our results, previous research found that tumor necrosis factor has a positive influence in the male gonadal tract with an impact on peritubular cell secretion and sperm survival (Puchner et al. 2012). Furthermore, several metabolism-related GO terms, including cholesterol metabolism process and secondary alcohol metabolism process, were enriched for KS (Fig. 6A). Indeed, KS men have an increased vulnerability to develop metabolic disease such as dyslipidaemia (Salzano et al. 2016). Previous study demonstrated altered fatty acid metabolism in adolescent patients with KS (Davis et al. 2023). In addition to the metabolism-related alteration, our results also indicated that PI3K/AKT pathway was enriched for KS. This result can be potentially explained by the essential role of AKT/mTOR pathway in mediating primordial germ cell growth, spermatogonial proliferation, and spermatogenesis (Miguel et al. 2002). In particular, mTOR play critical roles in spermatogenesis through mediating self-renewal and differentiation of spermatogonial stem cell, as well as the metabolic processes of Sertoli cells (Guan et al. 2020, Ma et al. 2016).

In patients with KS, the activation of immune status is found both in the testis and in the peripheral blood (Zhao, Yao, Xing, Jing, Li, Zhu, Yang, Zhai, Tian, Chen, Luo, Liu, Deng, Lin, Li, Fang, Sun, Wang, Zhou and Li 2020), and the impact of immune-related factors cannot be neglected for the pathogenesis of KS. Macrophages are the largest population of testicular immune cells. One recent study reported that macrophage polarization in the testis plays important roles in the onset of non-obstructive azoospermia (Zheng et al. 2021). Macrophages are heterogeneous immune cells and can be roughly divided into two distinct types, namely M1 and M2 macrophages. M1 macrophages protect against pathogenic microorganisms through stimulating immune response, whereas M2 macrophages have anti-inflammatory functions through inflammation resolving and wound healing (Martinez and Gordon 2014). Our results displayed the significant correlation between BPI and macrophages in KS patients (Fig. 7C). Human NK cells can be categorized as 2 subtypes (CD56bright and CD56dim) in humans. CD56bright NK
cells in the tissue were demonstrated to be detected within the lymphoid tissues, liver, and uterus (Melsen et al. 2016). Our results observed significant associations between ITLN1 and XG and CD56bright NK cells. Previous research revealed that dendritic cell abundance were negatively associated with semen quality (Duan et al. 2014), raising the possibility that dendritic cell-derived immune response could lead to male infertility. Similar to these previous reports, we found that BPI expression was significantly correlated with dendritic cells in KS patients (Fig. 7C). Furthermore, scRNA-seq data analysis showed strong correlations between macrophages and the other immune cell types, implying a high ligand-receptor interaction intensity mediated by macrophages.

In conclusion, we detected hub genes, critical pathways and infiltrating immune cells involved in the onset of KS. These results may deepen our understanding of the molecular mechanism of KS pathogenesis.

**Declarations**

**Ethical approval**

The medical research ethics committee of Nanjing Municipal Center for Disease Control and Prevention approved study design of this work.

**Consent for publication**

All authors.

**Availability of supporting data**

The dataset in this study can be obtained from the GEO website.

**Competing interests**

All co-authors declared no conflict of interest related to this study.

**Funding**

This study was supported by Jiangsu Funding Program for Excellent Postdoctoral Talent (2022ZB730)

**Authors’ contributions**


**Acknowledgements**

We would like to thank Yali Zhou from Department of Microbiology, Faculty of Basic Medical Sciences, Guilin Medical University for the proof reading of our manuscript.
Authors' information

L.L.T., Y.Y., Z.Q.M., D.D.X., Z.H.B. and M.K.Q. are from Nanjing Municipal Center for Disease Control and Prevention. T.C. is from Department of Urology, The First Affiliated Hospital of Nanjing Medical University. In addition, W.L. is from Center for Reproductive Medicine, Cheeloo College of Medicine, Shandong University.

References


Figures
**Figure 1**

Data preprocessing and differential expression analysis of GSE42331 and GSE47584. Boxplots of GEO datasets samples (A) before and (B) after normalization; Principal component analysis showing collective profile of the bulk datasets (C) prior to and (D) following batch effect elimination; (E) a volcano plot showing the DEGs between control and KS groups. DEG: differentially expressed gene; KS: Klinefelter Syndrome.
Figure 2

ROC curve analysis of the hub genes. ROC curve analysis of (A) DEFA4, (B) BPI, (C) MPO, (D) ITLN1 and (E) XG. ROC: receiver operating characteristic.

Figure 3
GSEA uncovering enriched hallmark pathways in KS. (A) A ridgeplot showing significantly enriched hallmark pathways for KS; (B) A cnetplot indicating gene symbols specific to top four enriched hallmark pathways; (C) A gseaplot demonstrating the top five significant hallmark pathways. GSEA, gene set enrichment analysis.

Figure 4

Enriched hallmark pathways for KS using GSVA. (A) A heatmap showing the correlations between each hallmark pathway and each control or KS sample; (B) a lollipop plot demonstrating significantly enriched hallmark pathway for KS. GSVA, gene set variation analysis.
Figure 5

Intersected enriched hallmark pathways for KS obtained by GSEA and GSVA.
GO and KEGG analysis using gseGO and gseKEGG functions. Bubble plots indicating the enrichment for (A) BP, (C) CC, and (E) MF GO terms and (G) KEGG pathways for KS; The upset plots showing enriched (B) BP, (D) CC, and (E) MF GO terms and (H) KEGG pathways for KS. BP: biological process; CC: cellular component; MF: molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes.
Figure 7

Infiltrated levels of immune cells in control and KS samples. (A) The heatmaps visualizing the association of each immune cell type and each control (left) or KS (right) sample; (B) Heatmaps of association visualizing the association of each immune cell type and each control (left) or KS (right) sample; (C) Bubble plots visualizing the significant correlations of immune cells with ITLN1, BPI and XG.
Figure 8

Cell-cell communication in KS patient based on scRNA-seq data. (A) Cluster analyses showing 10 clusters of captured cells through the UMAP method; (B) Identification of five cell types through manual annotation of the cell clusters (left) and a dotplot indicating expression distributions of cell type-specific marker genes (right); (C) The overall cell interaction network and each cell type-specific interaction network. The thicker the line represented, the more the number of interactions, and the stronger the interaction weights/strength between the two cell types. UMAP: uniform manifold approximation and projection.