Extrachromosomal Circular DNAs are Common and Functional in Esophageal Squamous Cell Carcinoma

Zhenguo Sun (sunzg@sdu.edu.cn)
Qilu Hospital, Cheeloo College of Medicine, Shandong University Qingdao https://orcid.org/0000-0002-3485-4034

Na Ji
Medical Division, Shandong Provincial Western Hospital

Renchang Zhao
Department of Thoracic Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University

Jinghui Liang
Department of Thoracic Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University

Jin Jiang
Department of Thoracic Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University

Hui Tian
Department of Thoracic Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University

Research

Keywords: eccDNAs, microDNA, ESCC, tumor progression, tumor induction, High-throughput sequencing, differential expression

Posted Date: November 12th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-103658/v1

Version of Record: A version of this preprint was published at Annals of Translational Medicine on January 1st, 2021. See the published version at https://doi.org/10.21037/atm-21-4372.

Abstract

Background

Esophageal squamous cell carcinoma (ESCC) is the leading cause of cancer-related mortality worldwide. Extrachromosomal circular DNAs (eccDNAs), especially the more-spread smaller type, were rarely studied before. Recently, the distribution of eccDNAs was reported to be distinguished in different tumor cell lineage or different condition of lung cancer. However, there was no exploration of eccDNA distribution and function in ESCC.

Methods

The eccDNAs from 3 surgical matched ESCC tissues were extracted and amplified after removing linear DNA / mitochondrial circular DNA and rolling circle amplification. High-throughput eccDNA sequencing and subsequent bioinformatics analysis was performed to study the distribution pattern and the level of eccDNAs in these matched tissues. Gene ontology and KEGG pathway analysis based on the genes associated with the eccDNAs at differential level were performed. 5 up-regulated and 5 down-regulated candidate eccDNAs were selected to be validated by routine PCR, TOPO-TA Cloning and Sanger sequencing. The nucleotides flanking eccDNA junctions were analyzed to explore the mechanism of eccDNA formation.

Results

184557 eccDNAs were identified. The overall length distribution was from 33bp to 968842bp, with the peak at ~360bp. They were originated mainly from 5’-untranslated regions, 3’-untranslated region, and rarely from exons, introns, LINE or Alu repeat region. The distribution patterns of the eccDNAs, such as chromosome distribution, length distribution and genomic annotation, were similar between ESCC and matched normal epithelium. Nevertheless, 16031 eccDNAs were found to be at differential level between ESCC and matched normal epithelium, including 10126 up-regulated eccDNAs and 5905 down-regulated eccDNAs. GO analysis and KEGG pathway analysis showed enriched pathways in cancer, MAPK pathway, GTPase related activity and cytoskeleton function et al. PCR, TOPO-TA Cloning and Sanger sequencing validated the junctional sites of 5 up-regulated candidate eccDNAs and 4 other unexpected eccDNAs. Repeat nucleotide pattern between the position flanking the start site and that flanking the end site was detected.

Conclusion

This study firstly demonstrated genome-wide presence of eccDNAs, explored the eccDNAs at differential level and revealed potential mechanism of eccDNAs in ESCC, thus suggesting potential clinical utility in ESCC and increasing our understanding about genome plasticity.

Background

Esophageal cancer ranks seventh in the incidence of cancers and sixth in cancer-related mortality worldwide [1]. Esophageal squamous cell carcinoma (ESCC) accounts for 90% of esophageal cancer globally and remains the dominant pathological type of esophageal cancer in high-risk areas, such as China [2]. Despite of rapid progress in the diagnosis and treatment of ESCC, the overall 5-year survival rate remains poor [3]. To our knowledge, the underlying molecular mechanisms causing high incidence and poor prognosis of ESCC have not fully elucidated.

The majority of human cellular DNA is contained in 22 linear autosome pairs and a pair of sex-determining chromosomes. Extrachromosomal DNA indicates the DNA elements separated from the chromosomes. The major form of extrachromosomal DNA is extrachromosomal circular DNA (eccDNA) featured by closed circular structure [4]. EccDNA was found half a century ago since double minutes (DMs) were detected in malignant tumor specimens of children [5]. The size of eccDNAs varies from hundreds of base pairs (bp) to several mega bases (Mb), however, the majority of eccDNA were short from 200–400 bps, also called microDNA or small poly-disperse circular DNA (spcDNA) [4, 6]. The eccDNA with longer size, such as DM, participates in cancer induction by carrying oncogenes and promoting gene amplification [7, 8]. However, less is known about the function and mechanism of the more-spread smaller type of eccDNAs because they are too small to contain coding genes [9]. Recently, by high-throughput sequencing, size distribution of the eccDNAs was identified to be varied between maternal and fetal plasma [10]. The genomic annotation of the eccDNA sequences weakly distinguished prostate from ovarian cancer cells [11]. The eccDNAs identified in lung cancer samples were longer than that in matched normal tissue. The size of circulating eccDNA in plasma decreased after surgical resection of lung cancer than before surgery [9]. These studies suggest the distribution of eccDNA could be correlated to tumor cell lineage and the eccDNA may be associated with the formation of certain tumors. However, there wasn't the exploration of eccDNA distribution and function in ESCC. Here, we are the first to investigate the distribution, function and mechanism of eccDNAs in ESCC.

Methods

Tissue specimens

This study was approved by the Ethic Committee at Qilu Hospital, Cheeloo College of Medicine, Shandong University in accordance with the ethical guidelines of the Declaration of Helsinki. Written informed consent was obtained from 3 enrolled patients with ESCC undergoing the surgery in the Department of Thoracic Surgery, Cheeloo College of Medicine, Shandong University. 3 pairs of tumor tissues and matched non-tumor esophageal epithelium were collected from the specimens dissected from above patients in the surgery.

Tissue DNA preparation and eccDNA sequencing

High-throughput eccDNA sequencing and subsequent bioinformatics analysis was performed by CloudSeq Biotech Inc. (Shanghai, China). Specially, 3 pairs of tumor and normal epithelium tissues were suspended in L1 solution (Plasmid Mini AX; A&A Biotechnology) and supplemented with Proteinase K

(ThermoFisher) before incubation overnight at 50 °C with agitation. After lysis, samples were alkaline treated, followed by precipitation of proteins and separation of chromosomal DNA from circular DNA through an ion exchange membrane column (Plasmid Mini AX; A&A Biotechnology). Column-purified DNA was treated with FastDigest MssI (Thermo Scientifific) to remove mitochondrial circular DNA and incubated at 37 °C for 16 h. Remaining linear DNA was removed by exonuclease (Plasmid-Safe ATP-dependent DNase, Epicentre) at 37 °C in a heating block and enzyme reaction was carried out continuously for 1 week, adding additional ATP and DNase every 24h (30 units per day) according to the manufacturer's protocol (Plasmid-Safe ATP-dependent DNase, Epicentre). eccDNA-enriched samples was used as template for phi29 polymerase amplification reactions (REPLI-g Midi Kit) amplifying eccDNA at 30 °C for 2 days (46–48 h). Phi29-amplifified DNA was sheared by sonication (Bioruptor), and the fragmented DNA was subjected to library preparation with NEBNext® Ultra II DNA Library Prep Kit for Illumina (New England Biolabs). Sequencing was carried out on Illumina NovaSeq 6000 with 150bp paired end mode according to the manufacturer's instructions.

Sequencing analysis of eccDNA

Paired-end reads were harvested from Illumina NovaSeq 6000 sequencer, and were quality controlled by Q30. After 3' adaptor-trimming and low quality reads removing by cutadapt software (v1.9.1), the high quality clean reads were aligned to the reference genome (UCSC hg19) with bwa software v (v0.7.12). Then, circle-map software (v1.1.4) was used to detect eccDNA within all samples, and samtools (v0.2) software was used to get raw soft-clipped read counts of the break point. Then edgeR (v0.6.9) software was used to perform normalization and differentially expressed eccDNA filter by p-value < 0.05 and fold change (>2 or <1/2). Bedtools (v2.27.1) software was used to annotate the eccDNAs. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway enrichment analysis were performed based on the differentially expressed eccDNA-associated genes. IGV (v2.4.10) software was used for eccDNA visualization.

Motif analysis of eccDNA junctional sites

We analyzed the nucleotide pattern flanking eccDNA junctional sites to explore the generation mechanism of these eccDNAs. Specially, the nucleotide composition from 10 bp upstream to 10 bp downstream of the start and end position inferred from the reference genome (hg19) were analyzed for each eccDNA locus.

Validation of eccDNA by routine PCR, TOPO-TA Cloning and Sanger sequencing

5 up-regulated and 5 down-regulated eccDNAs in different genomic regions and different chromosomes were chosen for experimental validation. Briefly, DNA was extracted from 3 matched esophageal squamous cell carcinoma specimens and treated with FastDigest MssI and exonuclease to remove mitochondrial circular DNA and linear DNA as described above. Then rolling circle amplification as above was performed to increase the yield. PCR with primers in these products using Accurate Taq Master Mix (dye plus) (Accurate Biotechnology, China) was performed to assess the candidate eccDNA level. The

reaction conditions were 94°C for 30 seconds, 28 cycles at 98°C for 10 seconds, 55°C for 30 seconds, and 72°C for 1 minute, followed by final elongation at 72°C for 2 minutes and storage at 4°C. The primers of eccDNA were designed using the “out-facing” strategy and they were described in Supplementary Table S1. Polymerase chain reaction (PCR) products were loaded onto 1.5% agarose gels and visualized under an ultraviolet Luminescent Image Analyzer (LAS-4000 Mini; GE Healthcare Life Sciences, Pittsburgh, US). The PCR product with specific positive bands were purified and amplified with TOPO-TA Cloning (Zero TOPO-TA Cloning Kit, Shanghai Yeasan, China) and sent for Sanger sequencing (Shanghai Sangon Biotech, China). The comparation of the nucleotide composition of each positive PCR product between Sanger sequencing and high-throughput sequencing was performed.

Results

Genome-wide detection and analysis of eccDNAs in the matched ESCC tissue

After DNA isolation, removal of linear DNA and mitochondrial circular DNA, rolling amplification and high throughput sequencing, there were more than 100 millions of clean reads in each sample after removing low quality reads. By mapping these clean reads to human genome (UCSC hg19), 184557 eccDNAs annotated at 23 pair of chromosomes were identified in these specimens. Most of these eccDNAs were detected in more than one specimen (Supplementary Table S2). These results indicated the existence of eccDNAs was a common event in ESCC tissues.

The genomic distribution of eccDNAs revealed that they were common in each of 23 pairs of chromosomes. No eccDNAs from mitochondrion was detected because they have been removed before sequencing. The eccDNA frequency per Mb was even in each chromosome, except chromosome Y with much lower frequency of eccDNAs (Fig. 1a/b). No correlation between the ratio of coding genes/Mb and eccDNA/Mb in each chromosome was found ($p = 0.27$) (Fig. 1c).

We mapped all the eccDNAs to different classes of genomic regions. Normalized genomic coverage was defined as the percentage of eccDNA mapped to that class of genomic regions divided by the percentage of the genome covered by that class of genomic region [10]. We found the eccDNAs were originated mainly from 5’-untranslated regions (5’-UTR), 3’-untranslated region (3’-UTR), 2 kb upstream or downstream of genes, 2 kb upstream to 2 kb downstream of CpG island regions etc.. Meanwhile, they were rarely distributed in exons, introns, LINE or Alu repeat region (Fig. 1d).

As to the eccDNAs relative to the deprived genes, 13177 genes in the genome gave rise to all the eccDNAs. There were the most 192 eccDNAs annotated at the gene of LSAMP. More than 100 eccDNAs were annotated at 16 respective genes (Supplementary Table S3). Only one eccDNA was originated from 3208 genes (Fig. 1e).

The overall length distribution of eccDNAs was from 33 bp to 968842 bp, with the peak at ~ 360 bp. Meanwhile, there were 2 additional peaks at ~ 555 bp and ~ 736 bp (Fig. 4A). 95.0% (175363/184557 of eccDNA were shorter than 3000 bp and 86.1% (158850/184557) were shorter than 2000 bp (Fig f/g).

The comparation of the distribution pattern of the eccDNAs between ESCC and matched normal epithelium

The distribution features varied between ESCC samples and matched normal esophageal epithelium. As Venn Diagram showed about the total of 184557 eccDNAs, 65809 eccDNAs were only detected in ESCC samples, 4520 were only detected in normal esophageal epithelium, while 114228 eccDNAs were detected in both samples (Fig. 2a). However, the chromosome distribution and annotation of genomic elements of eccDNAs was similar in ESCC and matched normal epithelium (Fig. 2b/c).

Because the length distribution of eccDNAs could distinguish maternal from fetal plasma and lung cancer from normal lung tissues in previous reports [9, 10], we analyzed the length distribution of eccDNAs in ESCC and matched normal esophageal epithelium. We found the length distribution in either ESCC or normal esophageal epithelium had similar features, such as the location of peak and the span of the length of eccDNAs (Fig. 2d).

Identification of the eccDNAs at differential level between ESCC and matched normal epithelium

Because most of eccDNAs were shown to be detected in more than one specimen, we further compared the level of eccDNAs in ESCC to that in normal esophageal epithelium to investigate the function of eccDNAs in ESCC. According to the screening criteria (p value < 0.05 and $|LogFC| > 1$), a total of 16031 eccDNA was defined as candidate functional eccDNAs, including 10126 up-regulated eccDNAs and 5905 down-regulated eccDNAs. Most of these candidate eccDNAs were detected in either ESCC samples or normal esophageal epithelium, while only a small fraction of candidate eccDNAs were detected in both of them (Fig. 3a/b/c). These candidate eccDNAs could distinguish ESCC from normal esophageal epithelium and may participate in the origin and progression of ESCC.

The length distribution of these candidate eccDNAs were from 44 bp to 395264 bp, with peak at ~ 357 bp and two additional peak at ~ 549 bp and ~ 733 bp, respectively (Fig. 3d/e). By mapping these candidate eccDNAs to genomic elements, we found they were mainly from 5'-UTR and 3'-UTR, and rarely from exons, introns or repeat regions such as LINE and Alu (Fig. 3f). Specially, among 10126 up-regulated eccDNAs, 49.4% (5007/10126) eccDNAs were annotated in the region of genes, while 50.6% (5119/10126) eccDNAs were annotated in the intergenic region. These eccDNAs were originated from 3219 genes. The most eccDNAs (13) were deprived from AUTS2 gene. The first 10 genes giving rise to the eccDNAs were shown in Supplementary Table S4. Among 5905 down regulated eccDNAs, 48.4% (2859) and 51.6% (3046) were annotated in the region of genes and intergenic region, respectively. They were deprived from 2235 genes. The most eccDNAs (7) were from LSAMP, CSMD1 and BICD1. The first 10 genes giving rise to the eccDNAs were shown in Supplementary Table S4.

GO and KEGG pathway analysis based on the genes associated with the eccDNAs at differential level

To understand the functions of genes associated with the eccDNAs at differential level, GO analysis was performed. GO analysis of genes associated with respective down-regulated or up-regulated eccDNAs included identification of cellular components (Fig. 4a/d), molecular function (Fig. 4b/e), and biological processes (Fig. 4c/f). Although the dominant biological processes were related to neurons, GTPase related activity and cytoskeleton were the main component in the molecular function and cellular components. In addition, the functions of the genes associated with eccDNAs at differential level were characterized by KEGG pathway analysis (Fig. 4g/h). The differentially expressed eccDNA-associated mRNAs were dominantly associated with pathways in cancer, mitogen-activated protein kinase (MAPK) pathway, focal adhesion, Rap1 pathway et al.

Discussion

This study originally investigated the features of eccDNAs in paired ESCC samples. EccDNAs were amplified by rolling circle amplification after removing linear DNA and mitochondrial circular DNA to yield the target product and to avoid bias caused by other DNA amplification [9]. Subsequent high-throughput sequencing demonstrated the common presence of eccDNAs in paired ESCC samples. Although Møller previously showed eccDNAs mapped mainly to gene-rich chromosomes, here we found the eccDNAs could be mapped to any region of the human genome and the frequency of eccDNAs formation was not associated with the chromosomes or coding genes [6]. Interestingly, the eccDNAs were far less frequently mapped to Y chromosome than other chromosomes maybe due to the less genetic information or denser structure etc. Specially, by mapping the eccDNAs to various genomic regions, they were originated dominantly from regulatory region such as 5’-UTR, 3’-UTR, CpG island, upstream or downstream of genes and less frequently from exons, introns, Alu or LINE repeat region, which pattern is a little different from the distribution pattern in previous studies [10]. To explore the cause of the specific tendency, we made the motif analysis of flanking eccDNA junctions to study the mechanism of eccDNA formation. We found repeat nucleotide pattern between the position flanking the start site and that flanking the end site. Previous studies showed the structure of 2–15 bp direct repeats flanking the junction site or dual direct repeats flanking the start and end position possibly contributed to the formation of eccDNAs through the circularization mediated by homologous recombination, microhomology end joining, nonhomologous end joining etc [4, 10]. Our analysis and these studies demonstrated that kinds of mechanism related to direct repeats may participate in the formation of eccDNA, which is necessary to be further explored.

In agreement with previous reports, the length distribution of detected eccDNAs implied that most of them were less than 1 kb with 3 peaks at 360 bp, 555 bp and 736 bp. However, unlike the significantly different length distribution of eccDNAs between lung cancer samples and matched normal tissue, before and after surgery, or maternal and fetal plasma as previously showed [6, 9], there was no obvious difference between ESCC and matched normal epithelium. Additionally, we did not observe any significant difference in the eccDNAs with regards to chromosome distribution and specific genomic regions between them. Surprisingly, the level of eccDNAs was obviously varied between ESCC and matched

normal epithelium. Kumar previously speculated whether the abundance of eccDNAs is altered in patients with cancer is difficult to be observed because rolling circle amplification with random primers would lose a massive information related to the level of a specific eccDNA [9]. Nevertheless, in our study with similar method based on screening criteria, abundant significantly up-regulated or down-regulated candidate eccDNAs potentially functioning in ESCC were finally selected according to high-throughput sequencing results. It can be explained by the fact that most of these candidate eccDNAs were detected exclusively in ESCC or matched normal tissue and only a small fraction of them were detected in both, thus easily being made a contrast between them. Moreover, the existence and the level of 5 up-regulated candidate eccDNAs and 4 unexpected up-regulated eccDNAs were successfully validated using routine PCR and Sanger sequencing, verifying the authenticity of our high throughput sequencing results. Interestingly, the absence of 5 down-regulated eccDNAs in PCR possibly due to the relative low expression indicated by few split read counts in high-throughput sequencing suggests that high throughput sequencing is more efficient than routine PCR to evaluate those candidate eccDNAs at low expression. Nevertheless, in our opinion, maybe more efficient way of rolling circle amplification or more advanced techniques other than rolling circle amplification to massively increase the yield of products after removing mitochondrial circular DNA and linear DNA are necessary to be further studied and applied, so that PCR, as a convenient and economic method, could be widely used to study the function and mechanism of eccDNAs in future. It is noteworthy that the validation of 4 unexpected up-regulated eccDNAs reminded us of designing more specific primers in PCR to amplify the specific product, especially when there were two or more partially-overlapped eccDNAs in the same sample. It should be highly emphasized and paid much attention to when we further study the function and mechanism of one specific eccDNA in ESCC or other tumors. In addition, the presence of the candidate eccDNAs at altered level demonstrated the deletion or amplification of a specific candidate eccDNA caused by certain stimulus, maybe like gene mutation etc., in normal esophageal epithelium might lead to the tumorigenesis and progression of ESCC. However, the exact mechanism remains to be explained and studied.

We further explored the possible function and mechanism of the eccDNAs in ESCC. GO analysis and KEGG pathway analysis using the genes generating the candidate eccDNAs at differential level showed enriched pathways in cancer, MAPK pathway, focal adhesion pathway, Rap1 pathway, GTPase related activity and cytoskeleton function etc., all of which was previously known to play key roles in ESCC progression [12–15]. As to investigate how the eccDNAs affect those pathways or functions, most of eccDNAs were less than 1 kb, and thus is too small to be like longer DM carrying oncogenes to participate in cancer induction. Møller previous identified 25 transcripts across eccDNA junction by screening mRNA sequencing and speculated a fraction of eccDNAs could be transcribed [6]. Nevertheless, we did not observe any transcripts across eccDNA junction by screening our whole transcriptome sequencing results in the same 3 paired ESCC samples (data not shown). We think the transcription of eccDNAs across eccDNA junction is at least not a dominant mechanism in ESCC progression. Recently, Paulsen reported shorter eccDNAs (microDNA) carrying miRNA gene or sequences of exons could generate functional small regulatory RNA, such as microRNA (miRNA) or novel si-like RNA, independent of a canonical promoter sequence to modulate related gene expression in vitro and in vivo [16]. In our results shown

above, the candidate eccDNAs at differential level were mainly annotated at 5’-UTR, 3’-UTR and CpG region rather than exons, introns or repeat sequences. Previous studies demonstrated 5’-UTR, 3’-UTR and CpG region are playing important roles in regulation of RNA translatability and stability or RNA transcription [17–19], We think maybe miRNA or si-like RNA generated potentially by the eccDNAs could affect the level of the corresponding RNA, especially those oncogene or tumor suppressor gene, transcriptionally or posttranscriptionally, and participate in ESCC induction and progression. However, whether this hypothesis is the dominant mechanism of eccDNA in ESCC or whether there is some other possible mechanism remained to be elucidated in future.

Conclusions

We firstly demonstrated the genome-wide presence of eccDNAs, explored the distribution and different level of eccDNAs, and revealed the potential mechanism of eccDNAs in paired ESCC samples. However, large size of samples is needed to reach final conclusion. In addition, more interesting questions remain to be answered in future. For examples, what is the specific protein or regulatory mechanism involving the formation of eccDNAs? What is the exact mechanism leading to the different level of eccDNAs in ESCC and matched normal epithelium? Does the amount of eccDNAs participate in the tumor heterogeneity? Do the eccDNAs play important roles in the prognosis of ESCC patients? Could the eccDNAs be as the useful clinical biomarker based on the level of cell-free eccDNAs in plasma of ESCC? In future, we will continue to research on these questions to investigate its potential clinical utility in ESCC and to increase our understanding about genome plasticity.

Abbreviations

ESCC
esophageal squamous cell carcinoma
eccDNA
extrachromosomal circular DNA
DMs
double minutes
spcDNA
small poly-disperse circular DNA
GO
Gene ontology
KEGG
Kyoto Encyclopedia of Genes and Genomes
PCR
Polymerase Chain Reaction
5’-UTR
5’-untranslated region

3’-UTR
3’-untranslated region
MAPK
mitogen-activated protein kinase
miRNA
microRNA

Declarations

Ethics approval and consent to participate

This study was approved by the Ethic Committee at Qilu Hospital, Cheeloo College of Medicine, Shandong University in accordance with the ethical guidelines of the Declaration of Helsinki. Written informed consent was obtained from 3 enrolled patients with ESCC undergoing surgery.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no conflict of interest.

Funding

This work was supported by National Natural Science Foundation of China (Grant no. 81802397 and 81672292), Natural Science Foundation of Shandong Province, China (Grant no. ZR2017BH035), China Postdoctoral Science Foundation (Grant no. 2020M672073), Project funded by Jinan Science and Technology Bureau (Grant no. 2019GXRC051), Taishan Scholar Program of Shandong Province (Grant no. ts201712087), Key research and development program of Shandong province (Grant no: 2019GSF108072).

Authors' contributions

ZS contribute to the study concept and design. HT reviewed this work. ZS and NJ performed the data analysis. RZ, JL and JJ collected the samples and performed experimental validation. ZS wrote this work. All the authors contribute to the interpretation of the data and approval of this work.

Acknowledgements

This work thanks Shanghai CloudSeq Biotech Inc for the assistance in the analysis of high-throughput sequencing data.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
2. Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, Cunningham D. Oesophageal cancer. Nat Rev Dis Primers. 2017;3:17048.
3. Yang H, Liu H, Chen Y, Zhu C, Fang W, Yu Z, Mao W, Xiang J, Han Y, Chen Z, et al. Neoadjuvant chemoradiotherapy followed by surgery versus surgery alone for locally advanced squamous cell carcinoma of the esophagus (NEOCRTEC5010): a phase III multicenter, randomized, open-label clinical rrial. J Clin Oncol. 2018;36:2796–803.
4. Shibata Y, Kumar P, Layer R, Willcox S, Gagan JR, Griffith JD, Dutta A. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. Science. 2012;336:82–6.
5. Cox D, Yuncken C, Spriggs AI. Minute chromatin bodies in malignant tumours of childhood. Lancet. 1965;1:55–8.
6. Møller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P, Maretty L, Hansen AJ, Snyder MP, Pilegaard H, et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. Nat Commun. 2018;9:1069.
7. Vogt N, Lefèvre SH, Apiou F, Dutrillaux AM, Cör A, Leuraud P, Poupon MF, Dutrillaux B, Debatisse M, Malfoy B. Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas. Proc Natl Acad Sci U S A. 2004;101:11368–73.
8. Storlazzi CT, Lonoce A, Guastadisegni MC, Trombetta D, Addabbo PD, Daniele G, Abbate AL, Macchia G, Surace C, Kok K, et al. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. Genome Res. 2010;20:1198–206.
9. Kumar P, Dillon LW, Shibata Y, Jazaeri AA, Jones DR, Dutta A. Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. Mol Cancer Res. 2017;15:1197–205.
10. Sin TKS, Jiang P, Deng J, Ji L, Cheng SH, Dutta A, Leung TY, Chan KCA, Chiu RWK, Lo YMD. Identification and characterization of extrachromosomal circular DNA in maternal plasma. Proc Natl Acad Sci U S A. 2020;117:1658–65.
11. Dillon LW, Kumar P, Shibata Y, Wang YH, Willcox S, Griffith JD, Pommier Y, Takeda S, Dutta A. Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. Cell Rep. 2015;11:1749–59.

12. Chen J, Zhang W, Wang Y, Zhao D, Wu M, Fan J, Li J, Gong Y, Dan N, Yang D, et al. The diacylglycerol kinase α (DGKα)/Akt/NF-κB feedforward loop promotes esophageal squamous cell carcinoma (ESCC) progression via FAK-dependent and FAK-independent manner. Oncogene. 2019;38:2533–50.
13. Wang K, Li J, Guo H, Xu X, Xiong G, Guan X, Liu B, Li J, Chen X, Yang K, et al. MiR-196a binding-site SNP regulates RAP1A expression contributing to esophageal squamous cell carcinoma risk and metastasis. Carcinogenesis. 2012;33:2147–54.
14. Shaverdashvili K, Padlo J, Weinblatt D, Jia Y, Jiang W, Rao D, Laczkó D, Whelan KA, Lynch JP, Muir AB, et al. KLF4 activates NFκB signaling and esophageal epithelial inflammation via the Rho-related GTP-binding protein RHOF. PLoS One. 2019;14:e0215746.
15. Kajiwara C, Fumoto K, Kimura H, Nojima S, Asano K, Odagiri K, Yamasaki M, Hikita H, Takehara T, Doki Y, et al. p63-dependent dickkopf3 expression promotes esophageal cancer cell proliferation via CKAP4. Cancer Res. 2018;78:6107–20.
16. Paulsen T, Shibata Y, Kumar P, Dillon L, Dutta A. Small extrachromosomal circular DNAs, microDNA, produce short regulatory RNAs that suppress gene expression independent of canonical promoters. Nucleic Acids Res. 2019;47:4586–96.
17. Jia L, Mao Y, Ji Q, Dersh D, Yewdell JW, Qian SB. Decoding mRNA translatability and stability from the 5' UTR. Nat Struct Mol Biol. 2020;27:814–21.
18. Lee WH, Han MW, Kim SH, Seong D, An JH, Chang HW, Kim SY, Kim SW, Lee JC. Tristetraprolin posttranscriptionally downregulates TRAIL death receptors. Cells. 2020;9:E1851.
19. Wu W, Bhagat TD, Yang X, Song JH, Cheng Y, Agarwal R, Abraham JM, Ibrahim S, Bartenstein M, Hussain Z. Hypomethylation of noncoding DNA regions and overexpression of the long noncoding RNA, AFAP1-AS1, in Barrett's esophagus and esophageal adenocarcinoma. Gastroenterology. 2013;144:956–66.

Tables

Table 1. The comparation between split read counts from high-throughput sequencing and PCR results.

Selected candidate eccDNAs		Samples					
		C1	C2	C3	N1	N2	N3
O1	high-throughput sequencing	8	235	1360	0	0	0
	PCR results	-	+	+	-	-	-
O2	high-throughput sequencing	41	98	3	0	0	0
	PCR results	+	+	-	-	-	-
O3	high-throughput sequencing	24	1	62	0	0	0
	PCR results	+	-	+	-	-	-
O4	high-throughput sequencing	13	20	2	0	0	0
	PCR results	-	+	-	-	-	-
O5	high-throughput sequencing	1	484	482	0	0	0
	PCR results	-	+	+	-	-	-
S1	high-throughput sequencing	0	0	0	7	5	6
	PCR results	-	-	-	-	-	-
S2	high-throughput sequencing	0	0	0	10	6	4
	PCR results	-	-	-	-	-	-
S3	high-throughput sequencing	0	0	0	9	2	3
	PCR results	-	-	-	-	-	-
S4	high-throughput sequencing	0	0	0	14	4	1
	PCR results	-	-	-	-	-	-
S5	high-throughput sequencing	0	0	0	11	6	2
	PCR results	-	-	-	-	-	-

Note: Ox indicates candidate up-regulated eccDNA, while Sx indicates candidate down-regulated eccDNA in ESCC. In high-throughput sequencing, the split read counts was shown. In PCR results, '+' indicates positive bands and '-' indicates negative bands.

Table 2. High-throughput sequencing and PCR results of candidate and unexpected eccDNAs

Primer type	EccDNAs ID	Detection type	Samples					
			C1	C2	C3	N1	N2	N3
O1 primer	O1 (chr3:153320590-153321556)	High-throughput sequencing	8	235	1360	0	0	0
	chr3:153320515-153321586		6	233	1	0	1	0
	O1 (chr3:153320590-153321556)	PCR results	-	-	+	-	-	-
	chr3:153320515-153321586		-	+	-	-	-	-
O2 primer	O2 (chr14:65484268-65484970)	High-throughput sequencing	41	98	3	0	0	0
	chr14:65484250-65484750		38	0	3	0	0	0
	O2 (chr14:65484268-65484970)	PCR results	-	+	-	-	-	-
	chr14:65484250-65484750		+	-	-	-	-	-
O3 primer	O3 (chr1:201630725-201631049)	High-throughput sequencing	24	1	62	0	0	0
	chr1:201630702-201631065		1	0	61	0	0	0
	O3 (chr1:201630725-201631049)	PCR results	+	-	-	-	-	-
	chr1:201630702-201631065		-	-	+	-	-	-
O5 primer	O5 (chr3:137490587-137491130)	High-throughput sequencing	1	484	482	0	0	0
	chr3:137490571-137491158		1	471	0	0	1	1
	O5 (chr3:137490587-137491130)	PCR results	-	-	+	-	-	-
	chr3:137490571-137491158		-	+	-	-	-	-

Figures

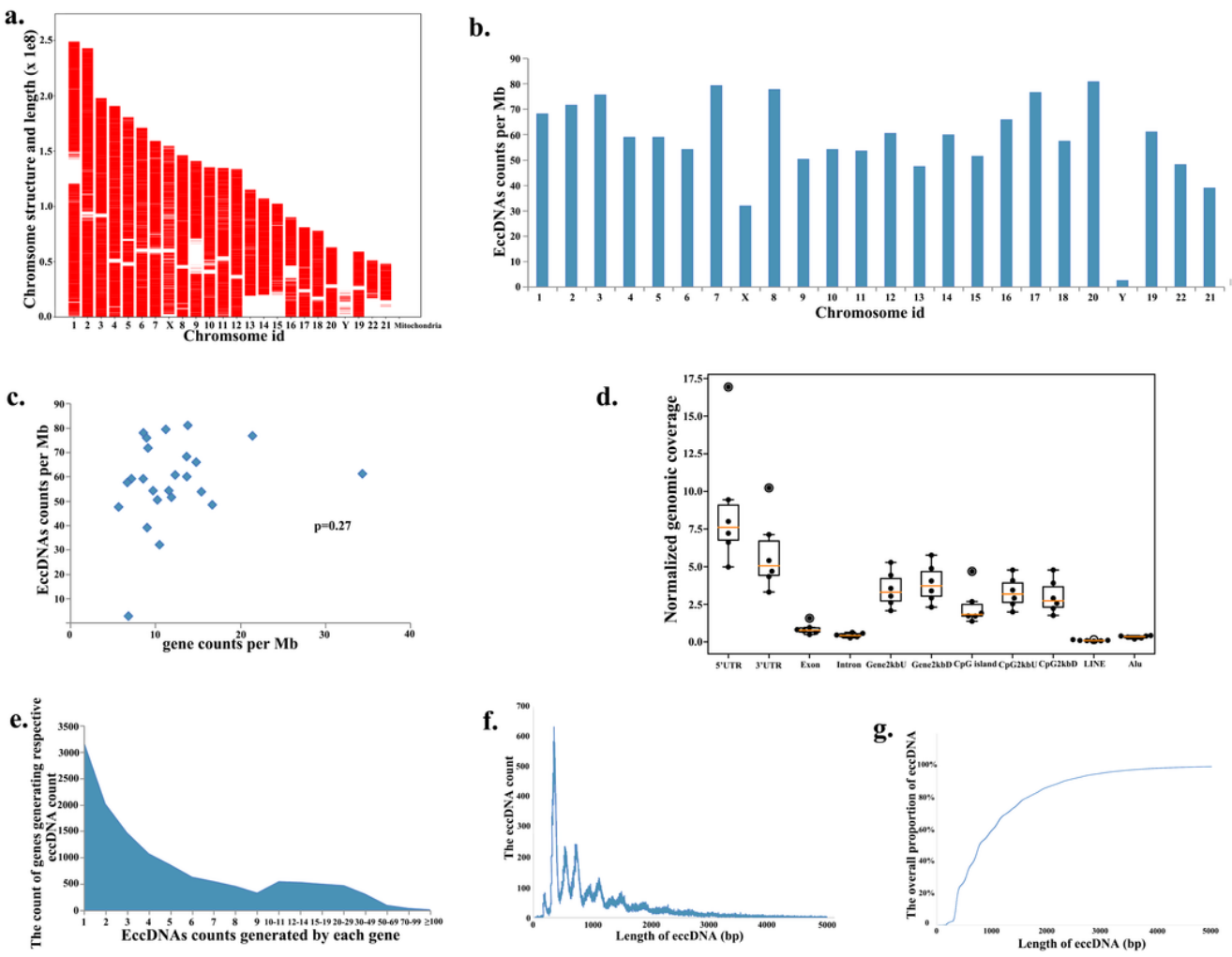

Figure 1

Genome-wide detection and analysis of eccDNAs distribution by high throughput sequencing in matched ESCC samples. a. The eccDNAs distribution in 23 pairs of chromosomes. b. The eccDNA frequency per Mb in each chromosome. c. The relationship between the ratio of coding genes/Mb and eccDNAs/Mb in each chromosome was not significant (p = 0.27). d. The distribution of eccDNAs in different classes of genomic regions. e. The distribution of the counts of eccDNAs generated by different genes. f/g. The length distribution of eccDNA at ≤5000bp.

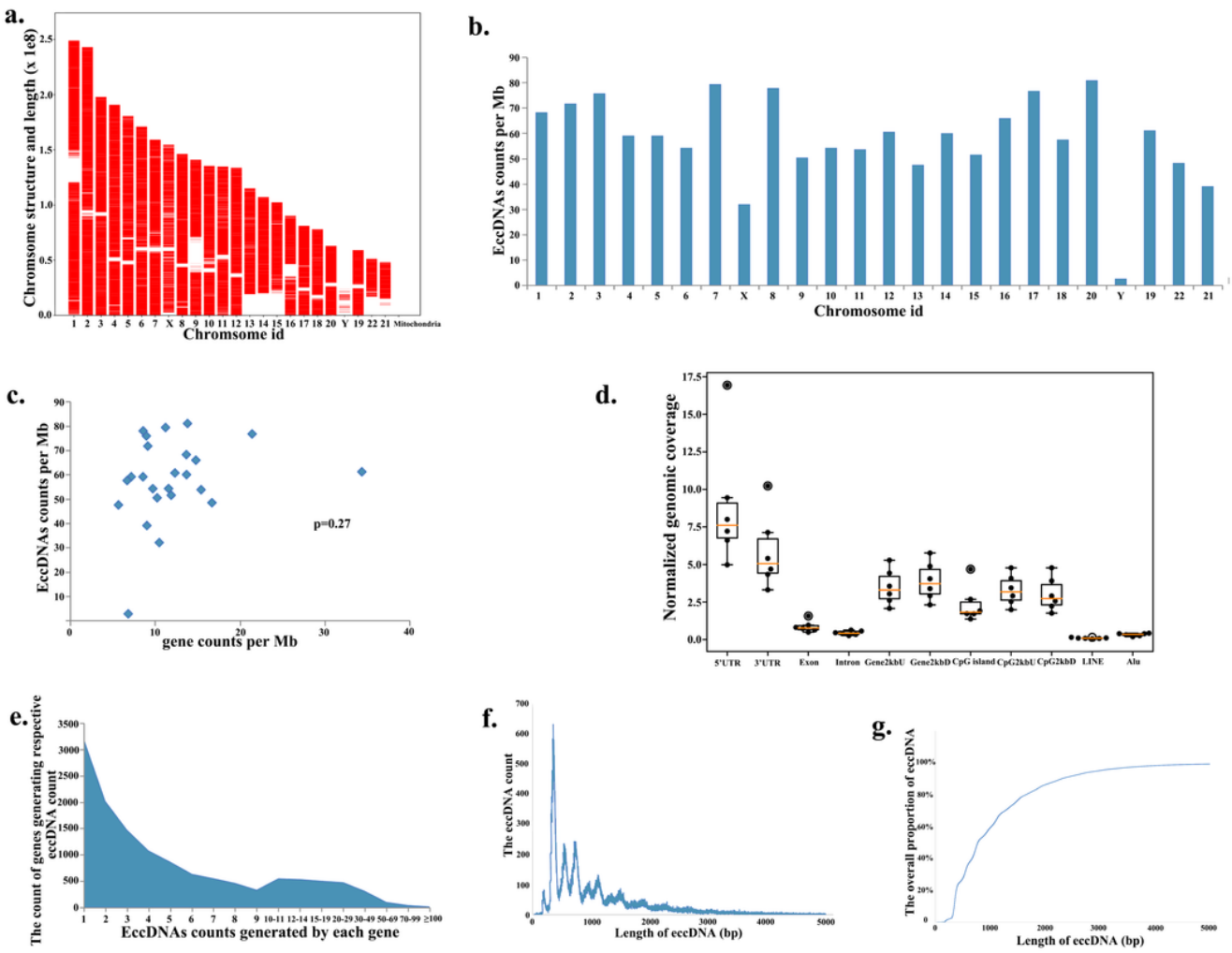

Figure 1

Genome-wide detection and analysis of eccDNAs distribution by high throughput sequencing in matched ESCC samples. a. The eccDNAs distribution in 23 pairs of chromosomes. b. The eccDNA frequency per Mb in each chromosome. c. The relationship between the ratio of coding genes/Mb and eccDNAs/Mb in each chromosome was not significant (p = 0.27). d. The distribution of eccDNAs in different classes of genomic regions. e. The distribution of the counts of eccDNAs generated by different genes. f/g. The length distribution of eccDNA at ≤5000bp.

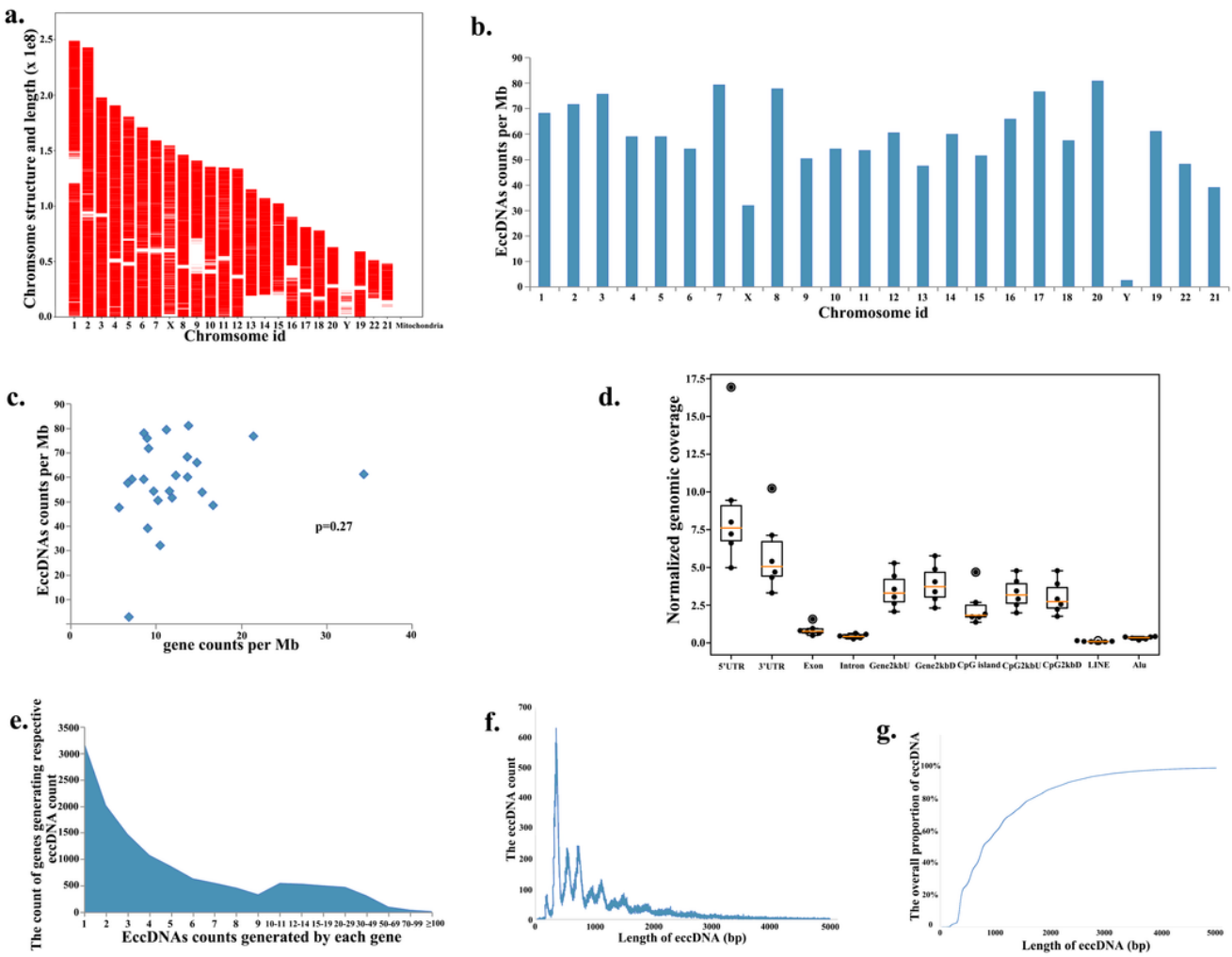

Figure 1

Genome-wide detection and analysis of eccDNAs distribution by high throughput sequencing in matched ESCC samples. a. The eccDNAs distribution in 23 pairs of chromosomes. b. The eccDNA frequency per Mb in each chromosome. c. The relationship between the ratio of coding genes/Mb and eccDNAs/Mb in each chromosome was not significant (p = 0.27). d. The distribution of eccDNAs in different classes of genomic regions. e. The distribution of the counts of eccDNAs generated by different genes. f/g. The length distribution of eccDNA at ≤5000bp.

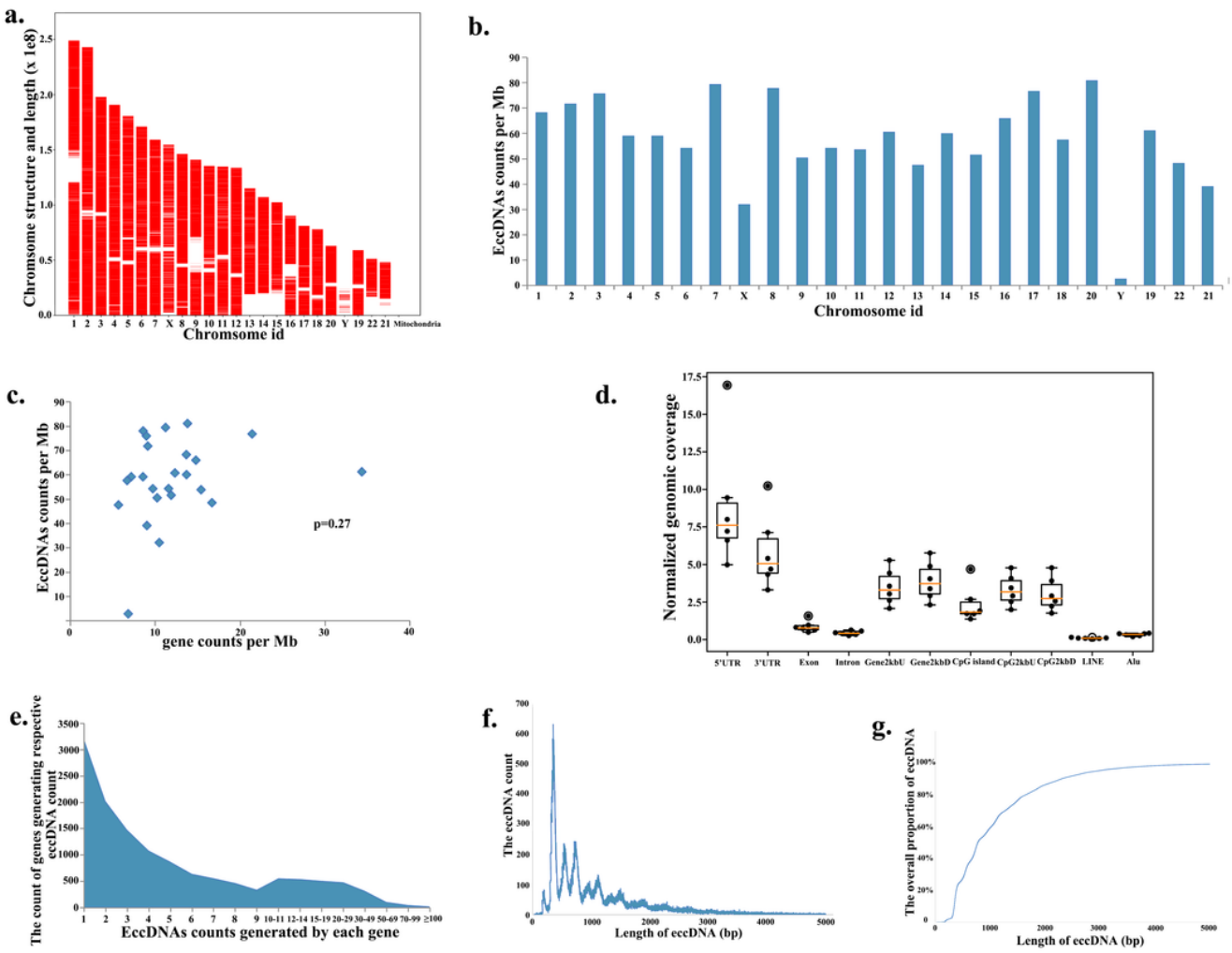

Figure 1

Genome-wide detection and analysis of eccDNAs distribution by high throughput sequencing in matched ESCC samples. a. The eccDNAs distribution in 23 pairs of chromosomes. b. The eccDNA frequency per Mb in each chromosome. c. The relationship between the ratio of coding genes/Mb and eccDNAs/Mb in each chromosome was not significant (p = 0.27). d. The distribution of eccDNAs in different classes of genomic regions. e. The distribution of the counts of eccDNAs generated by different genes. f/g. The length distribution of eccDNA at ≤5000bp.

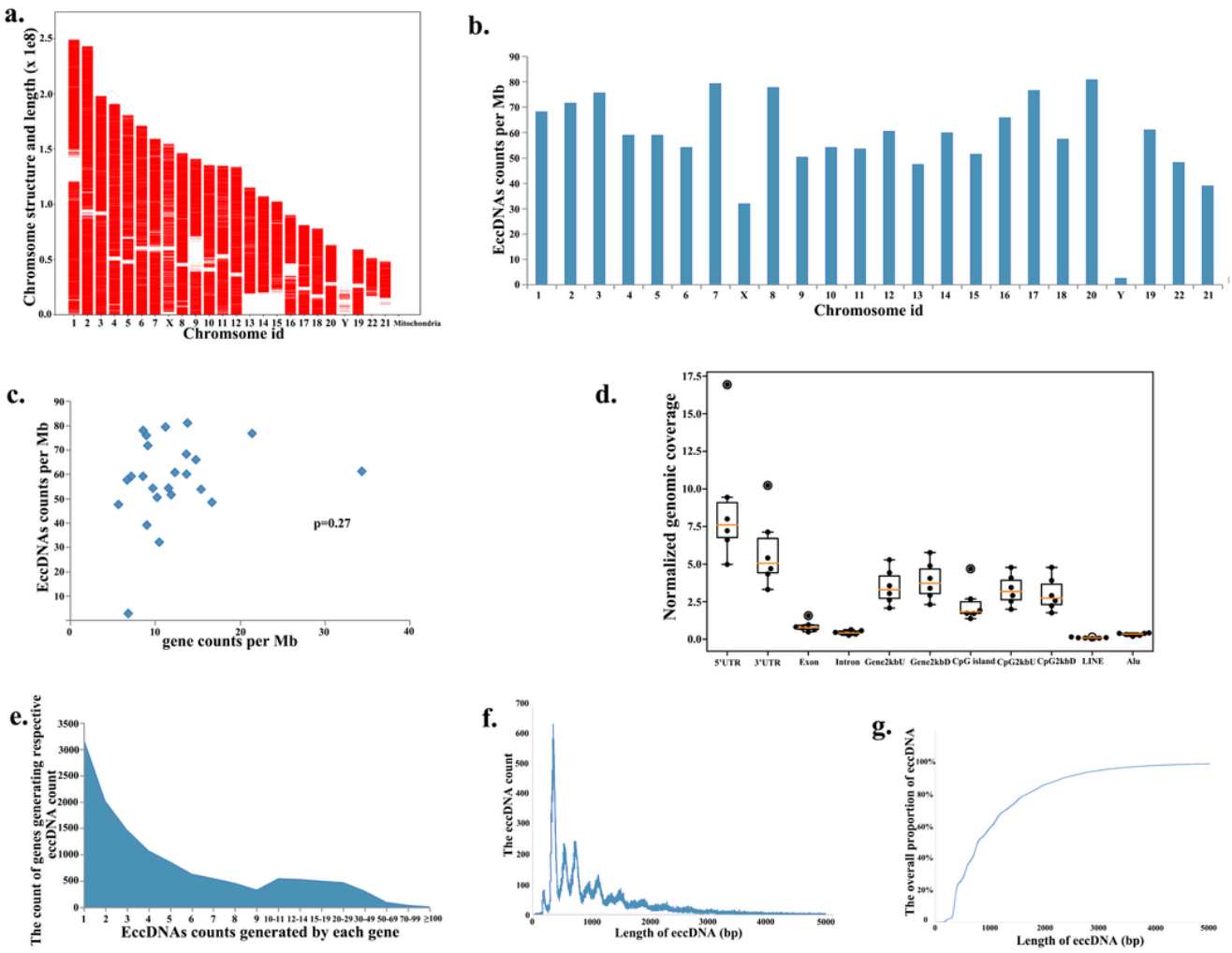

Figure 1

Genome-wide detection and analysis of eccDNAs distribution by high throughput sequencing in matched ESCC samples. a. The eccDNAs distribution in 23 pairs of chromosomes. b. The eccDNA frequency per Mb in each chromosome. c. The relationship between the ratio of coding genes/Mb and eccDNAs/Mb in each chromosome was not significant (p = 0.27). d. The distribution of eccDNAs in different classes of genomic regions. e. The distribution of the counts of eccDNAs generated by different genes. f/g. The length distribution of eccDNA at ≤5000bp.

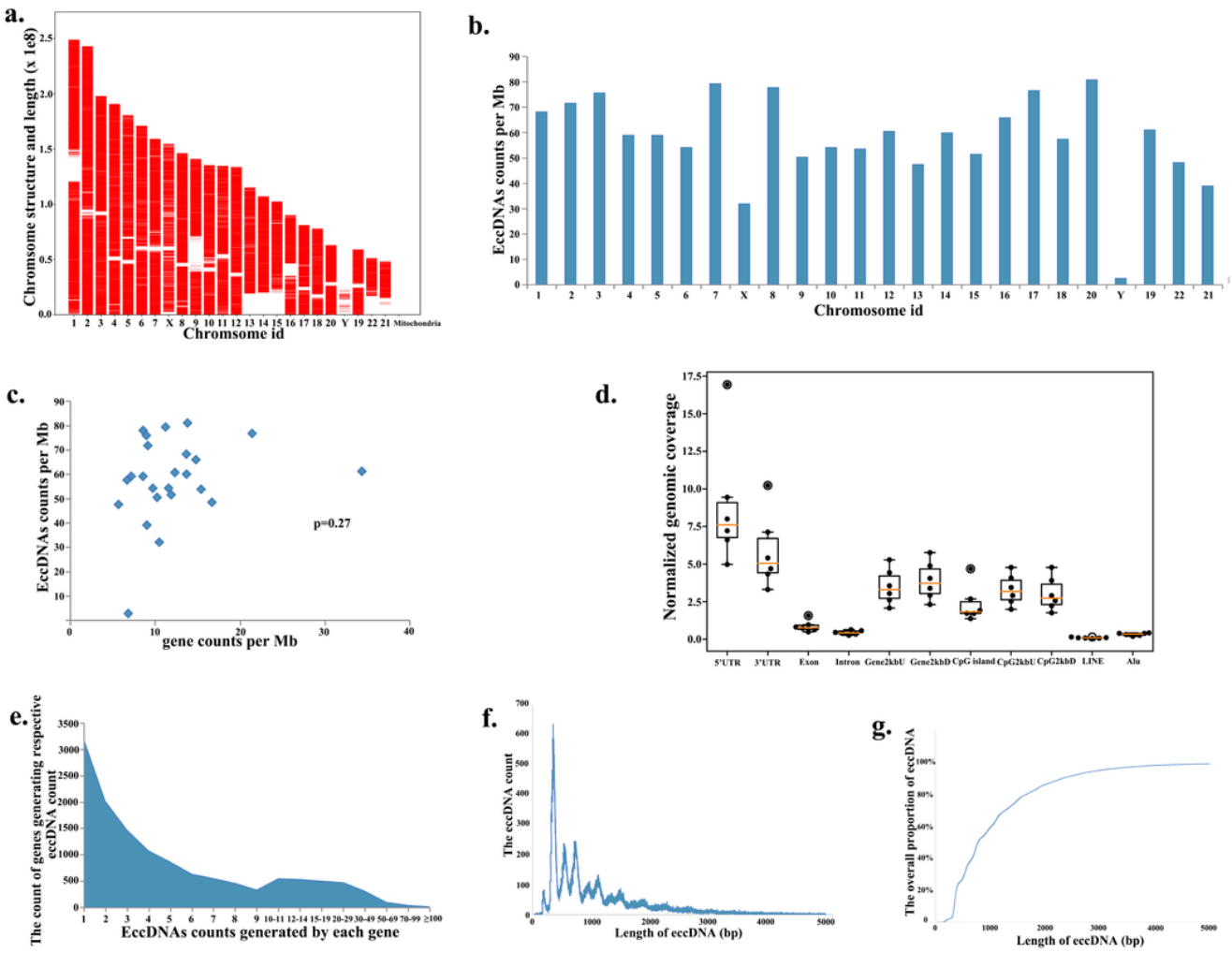

Figure 1

Genome-wide detection and analysis of eccDNAs distribution by high throughput sequencing in matched ESCC samples. a. The eccDNAs distribution in 23 pairs of chromosomes. b. The eccDNA frequency per Mb in each chromosome. c. The relationship between the ratio of coding genes/Mb and eccDNAs/Mb in each chromosome was not significant (p = 0.27). d. The distribution of eccDNAs in different classes of genomic regions. e. The distribution of the counts of eccDNAs generated by different genes. f/g. The length distribution of eccDNA at ≤5000bp.

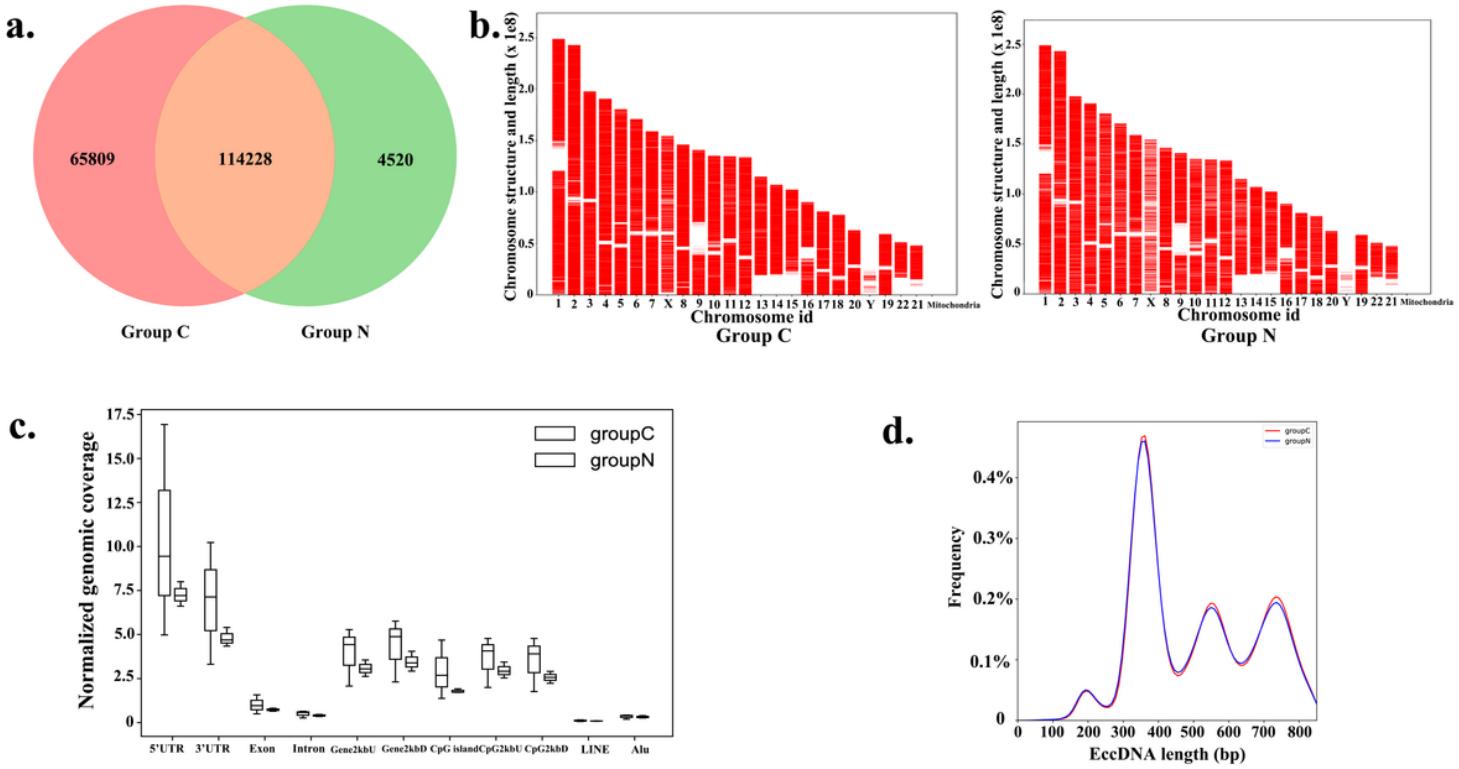

Figure 2

The comparation of the distribution pattern of the eccDNAs between ESCC and matched normal epithelium. a. Venn Diagram showing the eccDNAs detected in ESCC or matched normal epithelium. b/c/d. The chromosome distribution, annotation of genomic elements and length distribution of eccDNAs were similar in ESCC and matched normal epithelium.

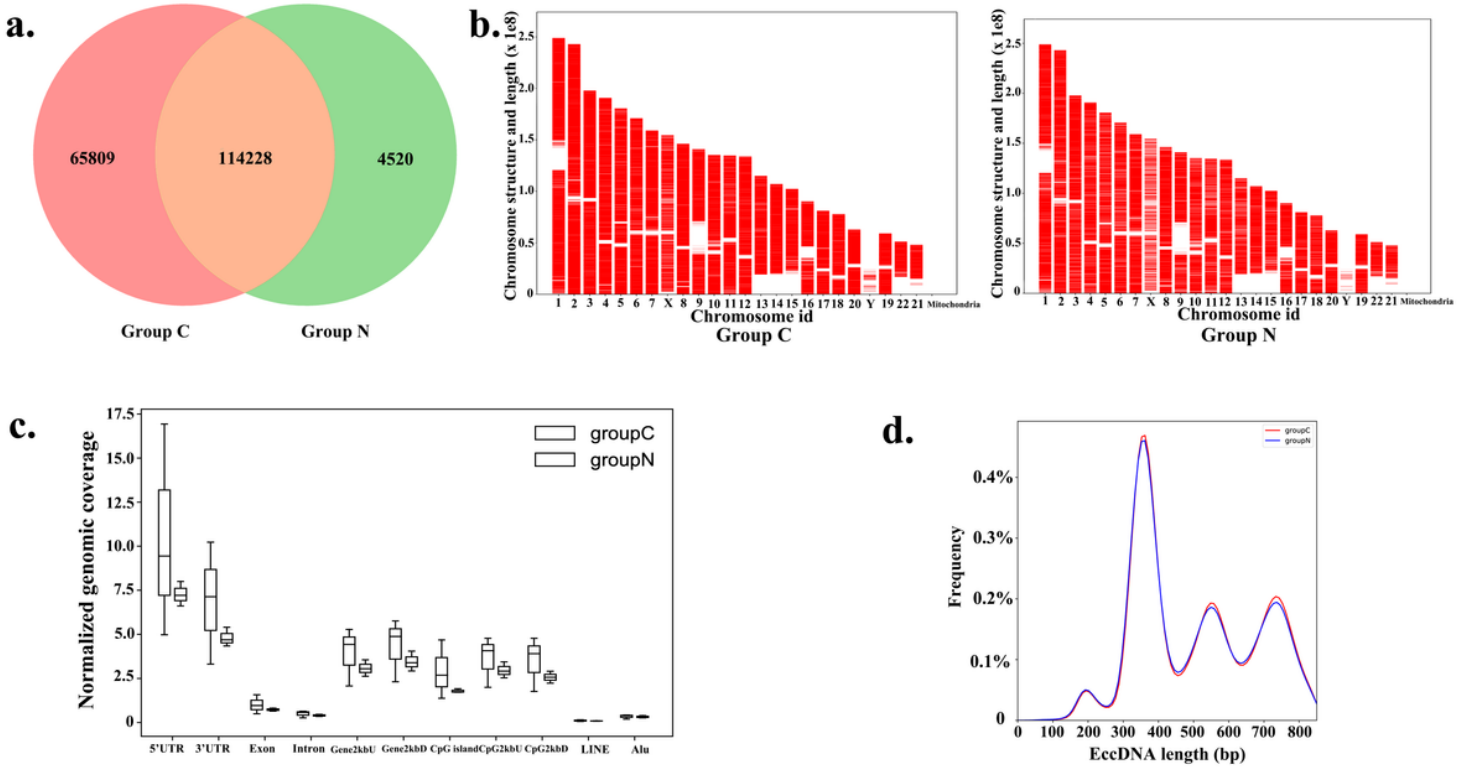

Figure 2

The comparation of the distribution pattern of the eccDNAs between ESCC and matched normal epithelium. a. Venn Diagram showing the eccDNAs detected in ESCC or matched normal epithelium. b/c/d. The chromosome distribution, annotation of genomic elements and length distribution of eccDNAs were similar in ESCC and matched normal epithelium.

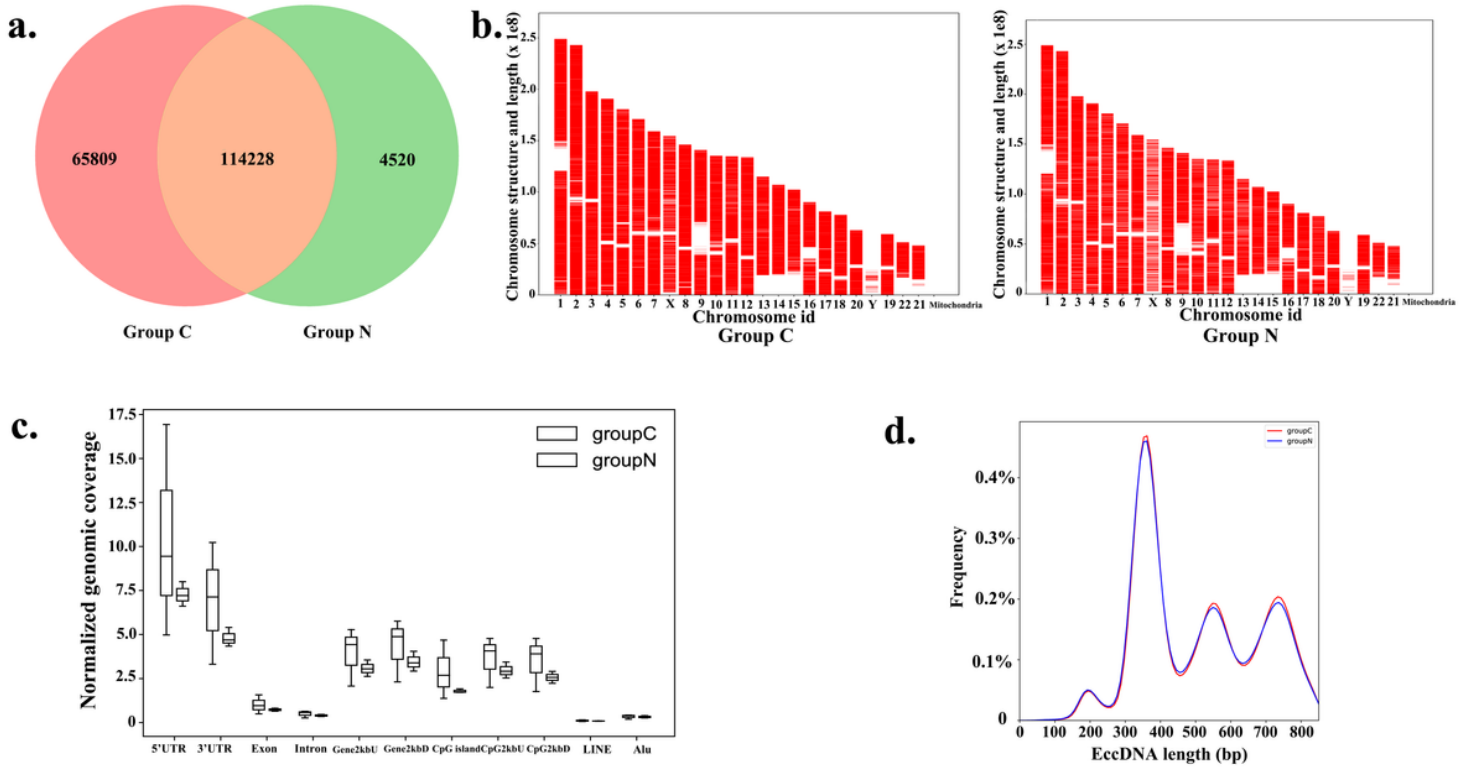

Figure 2

The comparation of the distribution pattern of the eccDNAs between ESCC and matched normal epithelium. a. Venn Diagram showing the eccDNAs detected in ESCC or matched normal epithelium. b/c/d. The chromosome distribution, annotation of genomic elements and length distribution of eccDNAs were similar in ESCC and matched normal epithelium.

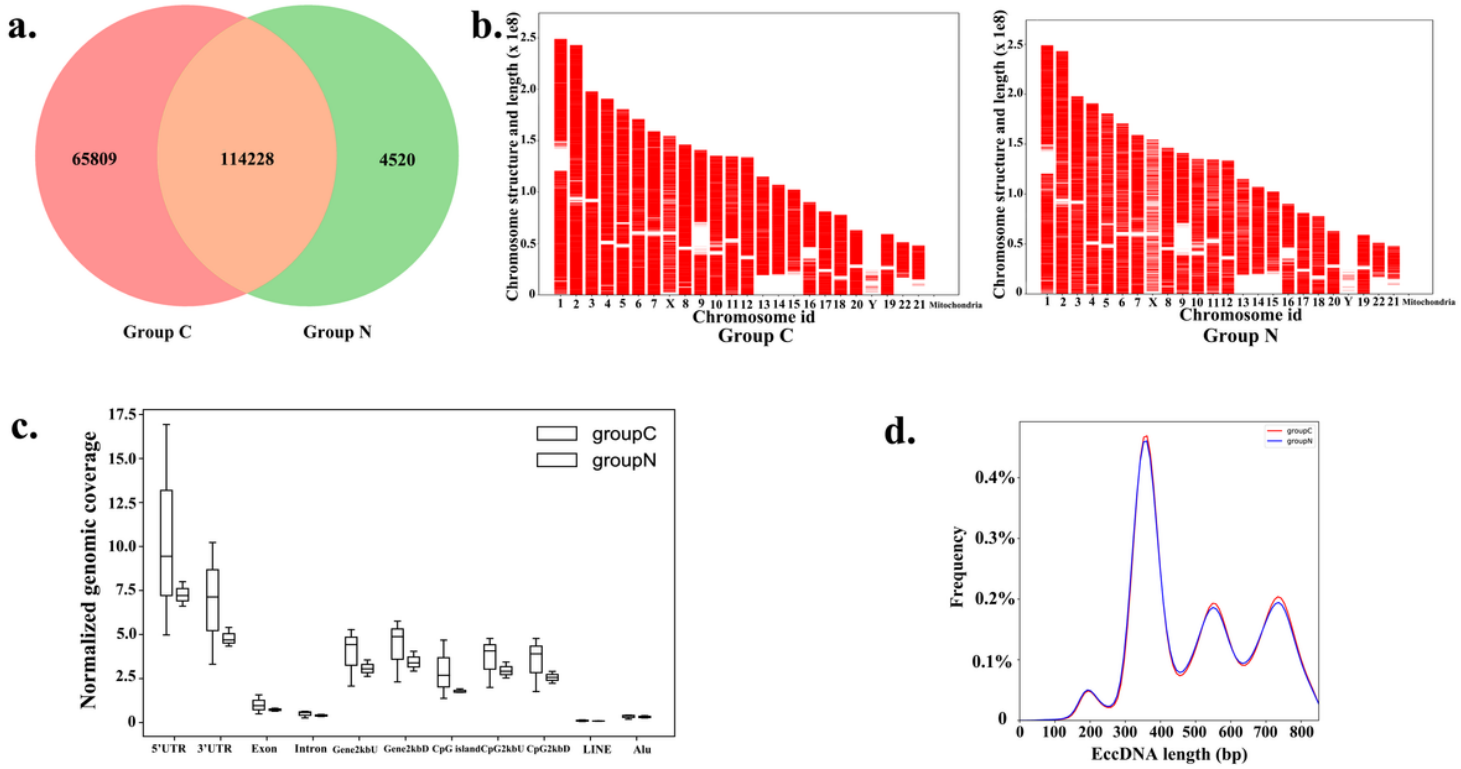

Figure 2

The comparation of the distribution pattern of the eccDNAs between ESCC and matched normal epithelium. a. Venn Diagram showing the eccDNAs detected in ESCC or matched normal epithelium. b/c/d. The chromosome distribution, annotation of genomic elements and length distribution of eccDNAs were similar in ESCC and matched normal epithelium.

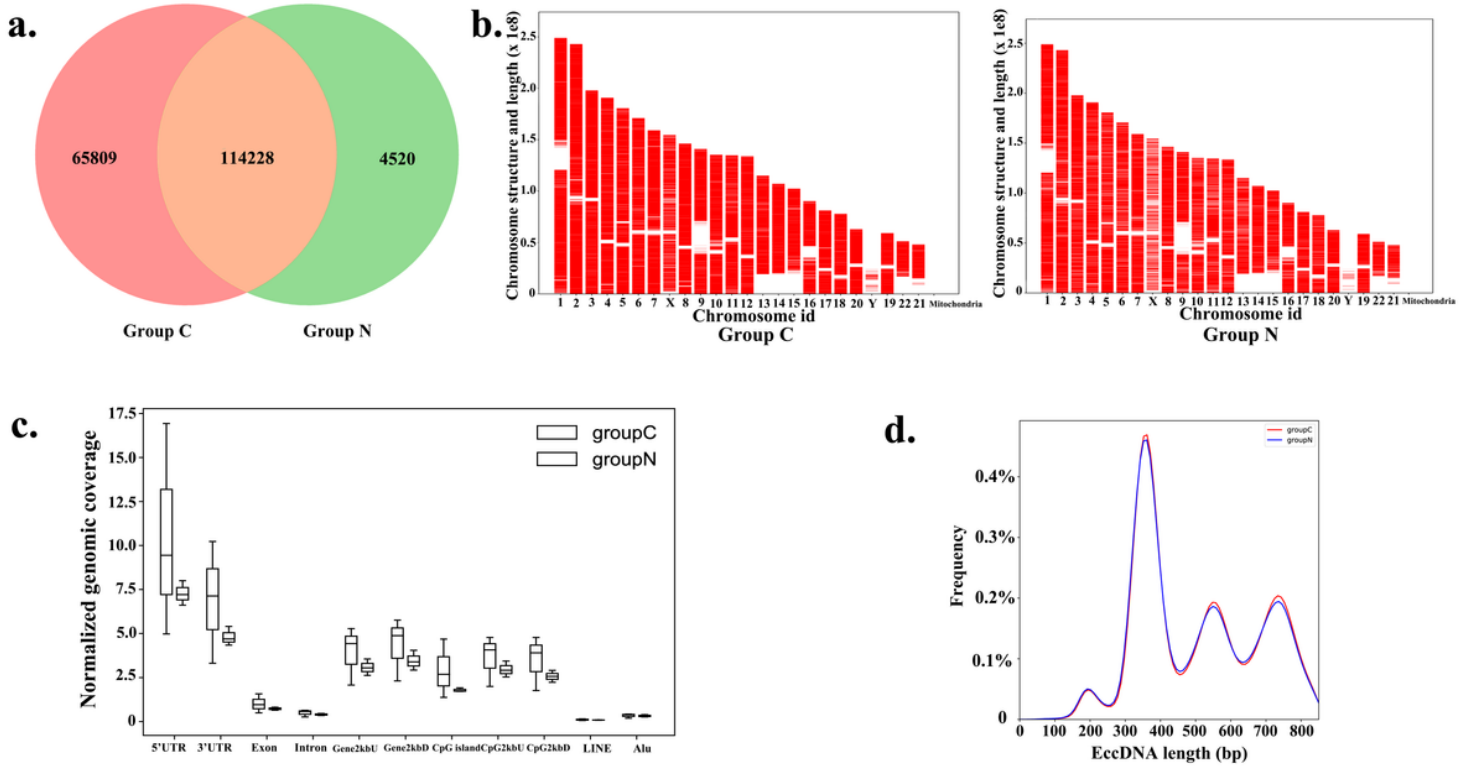

Figure 2

The comparation of the distribution pattern of the eccDNAs between ESCC and matched normal epithelium. a. Venn Diagram showing the eccDNAs detected in ESCC or matched normal epithelium. b/c/d. The chromosome distribution, annotation of genomic elements and length distribution of eccDNAs were similar in ESCC and matched normal epithelium.

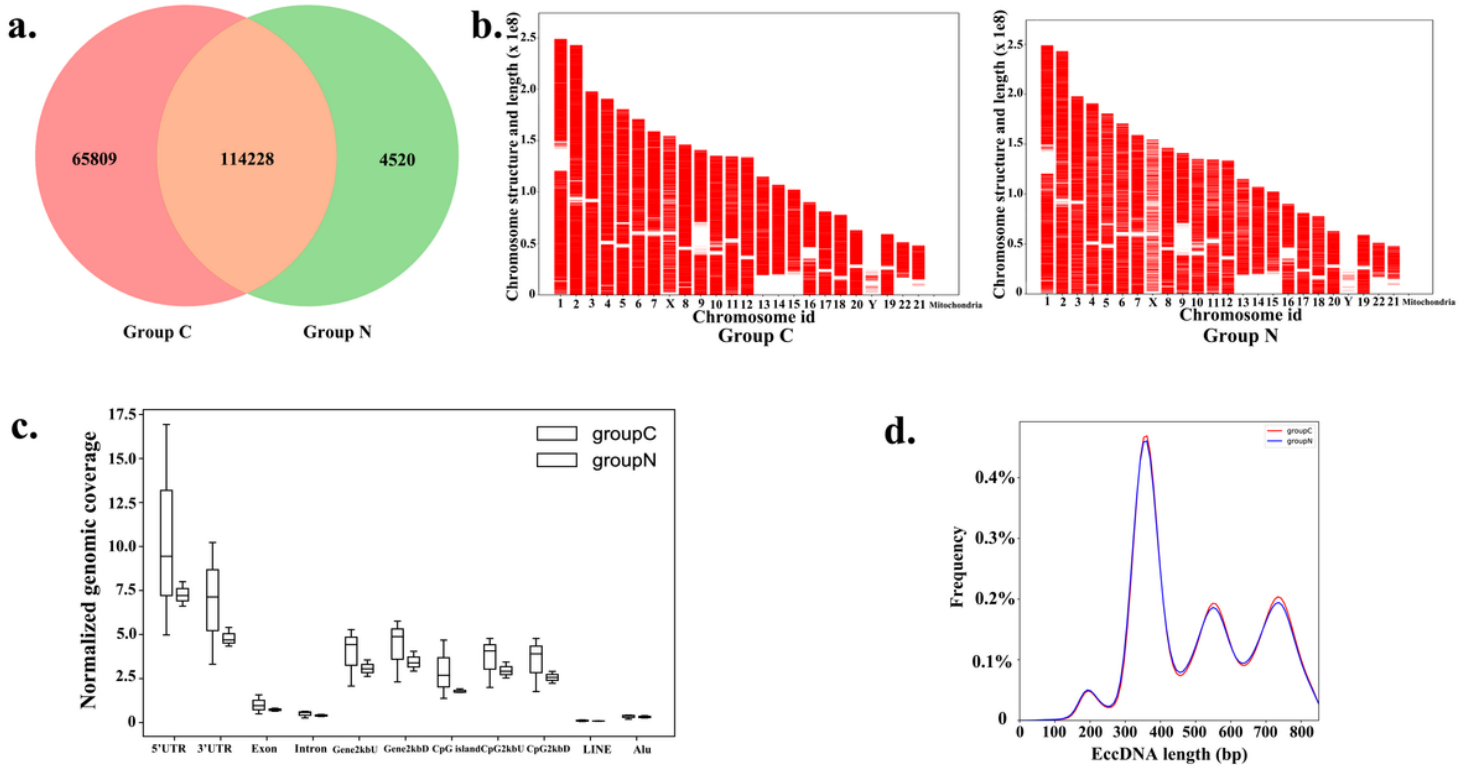

Figure 2

The comparation of the distribution pattern of the eccDNAs between ESCC and matched normal epithelium. a. Venn Diagram showing the eccDNAs detected in ESCC or matched normal epithelium. b/c/d. The chromosome distribution, annotation of genomic elements and length distribution of eccDNAs were similar in ESCC and matched normal epithelium.

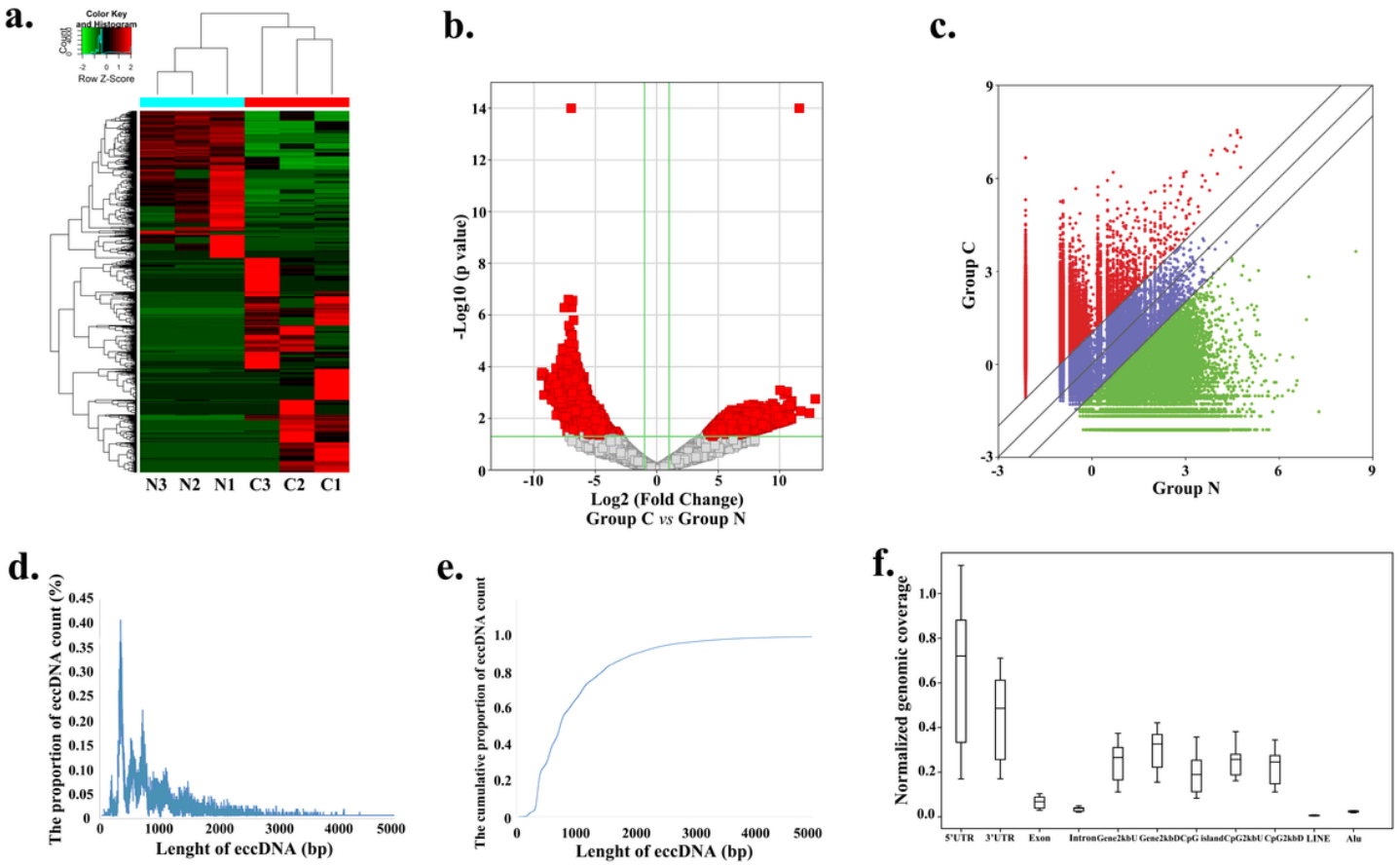

Figure 3

Identification of the eccDNAs at differential level between ESCC and matched normal epithelium. a/b/c. Cluster, scatter plot and volcano plot showing the candidate eccDNAs at differential level between ESCC and matched normal epithelium. d/e. Length distribution of the candidate eccDNAs at differential level. f. The distribution of the candidate eccDNA at differential level in different classes of genomic regions.

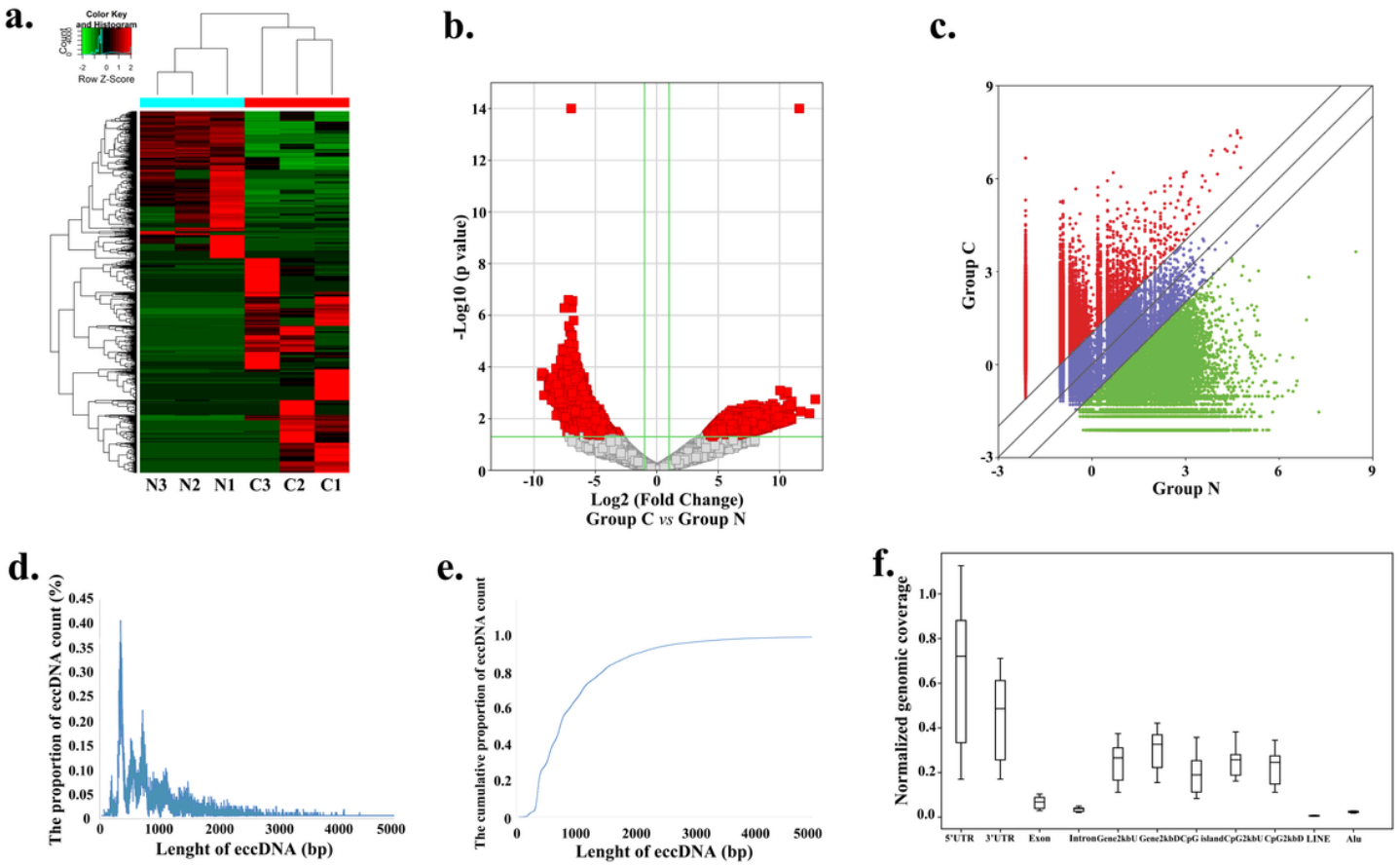

Figure 3

Identification of the eccDNAs at differential level between ESCC and matched normal epithelium. a/b/c. Cluster, scatter plot and volcano plot showing the candidate eccDNAs at differential level between ESCC and matched normal epithelium. d/e. Length distribution of the candidate eccDNAs at differential level. f. The distribution of the candidate eccDNA at differential level in different classes of genomic regions.

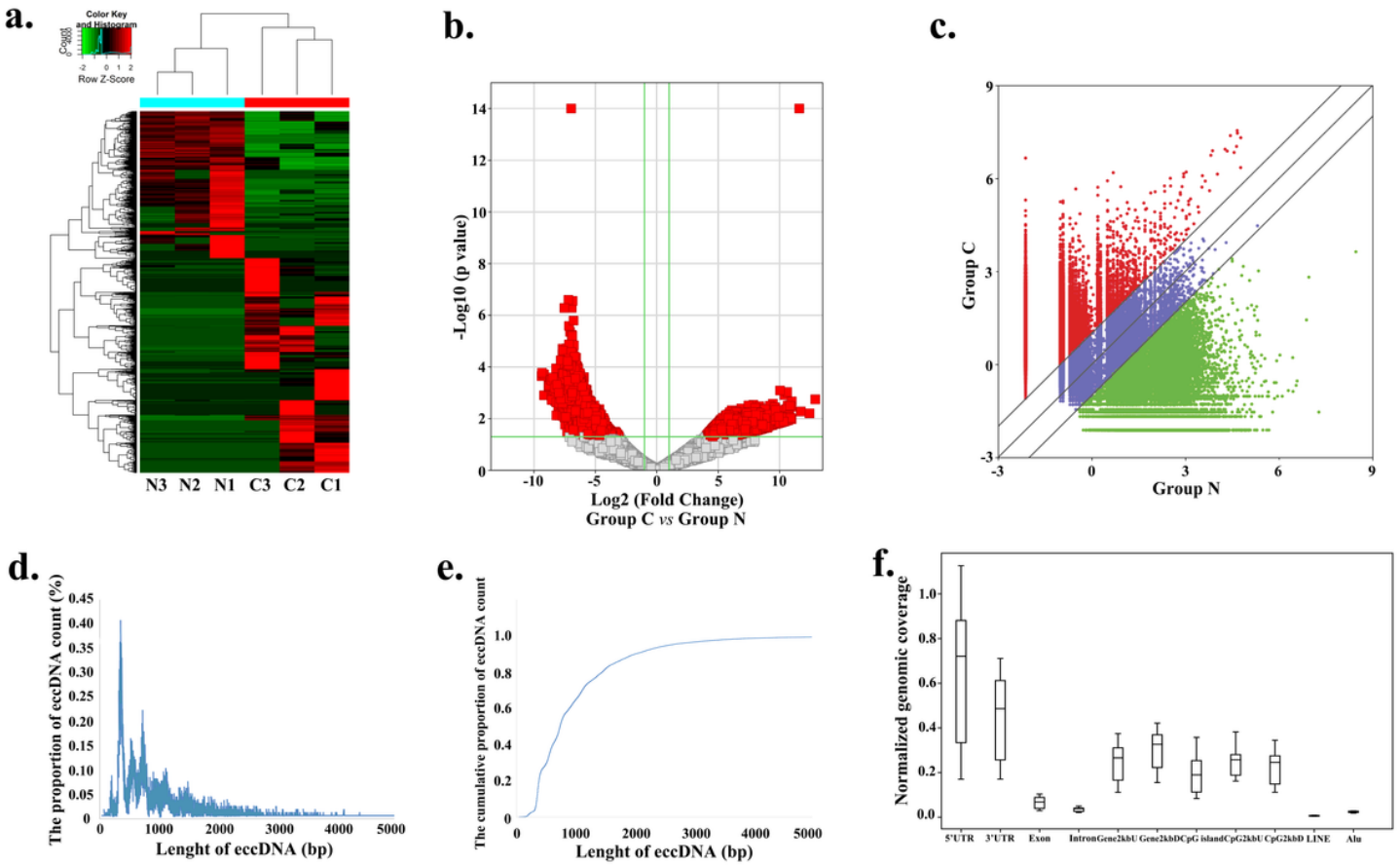

Figure 3

Identification of the eccDNAs at differential level between ESCC and matched normal epithelium. a/b/c. Cluster, scatter plot and volcano plot showing the candidate eccDNAs at differential level between ESCC and matched normal epithelium. d/e. Length distribution of the candidate eccDNAs at differential level. f. The distribution of the candidate eccDNA at differential level in different classes of genomic regions.

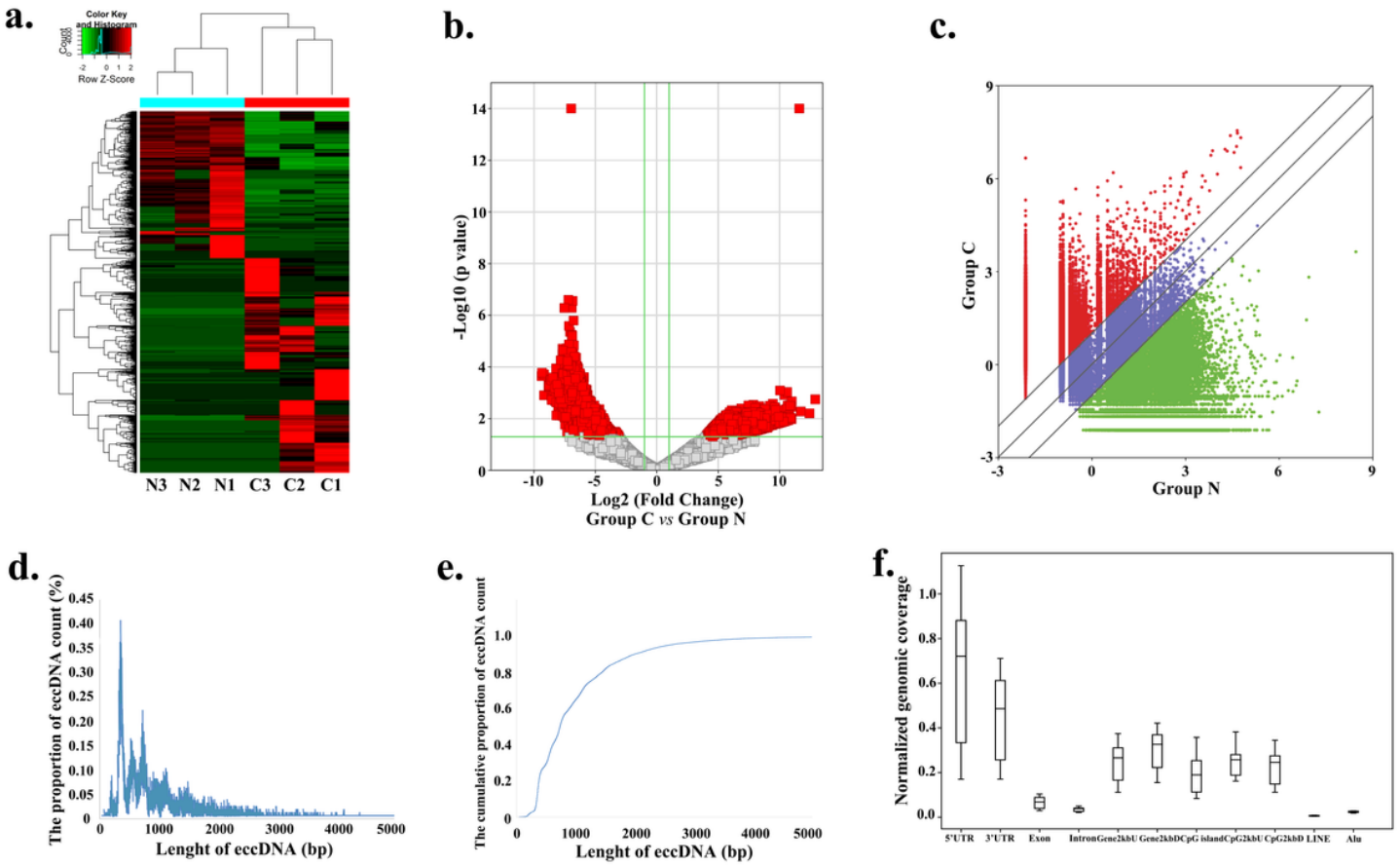

Figure 3

Identification of the eccDNAs at differential level between ESCC and matched normal epithelium. a/b/c. Cluster, scatter plot and volcano plot showing the candidate eccDNAs at differential level between ESCC and matched normal epithelium. d/e. Length distribution of the candidate eccDNAs at differential level. f. The distribution of the candidate eccDNA at differential level in different classes of genomic regions.

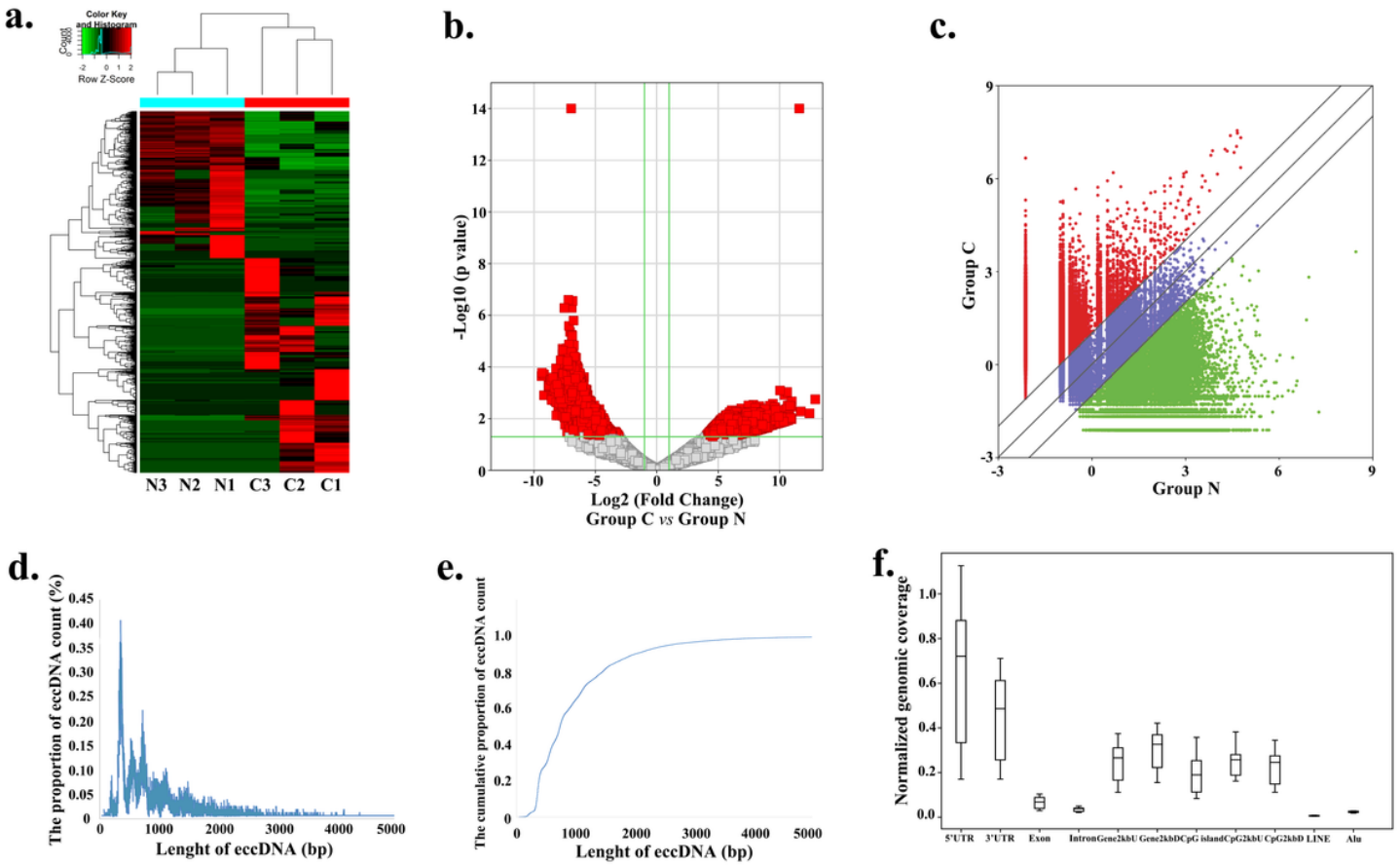

Figure 3

Identification of the eccDNAs at differential level between ESCC and matched normal epithelium. a/b/c. Cluster, scatter plot and volcano plot showing the candidate eccDNAs at differential level between ESCC and matched normal epithelium. d/e. Length distribution of the candidate eccDNAs at differential level. f. The distribution of the candidate eccDNA at differential level in different classes of genomic regions.

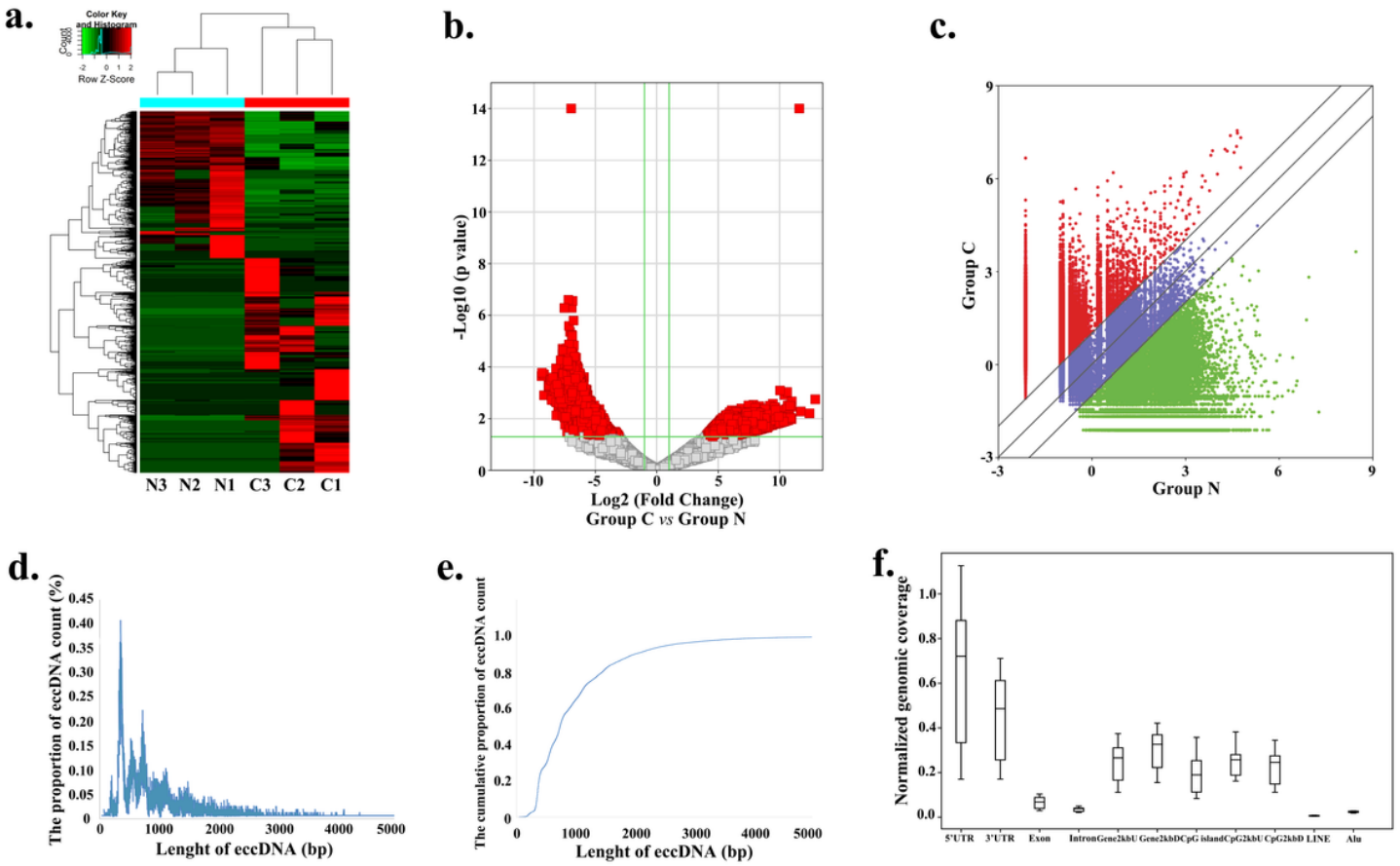

Figure 3

Identification of the eccDNAs at differential level between ESCC and matched normal epithelium. a/b/c. Cluster, scatter plot and volcano plot showing the candidate eccDNAs at differential level between ESCC and matched normal epithelium. d/e. Length distribution of the candidate eccDNAs at differential level. f. The distribution of the candidate eccDNA at differential level in different classes of genomic regions.

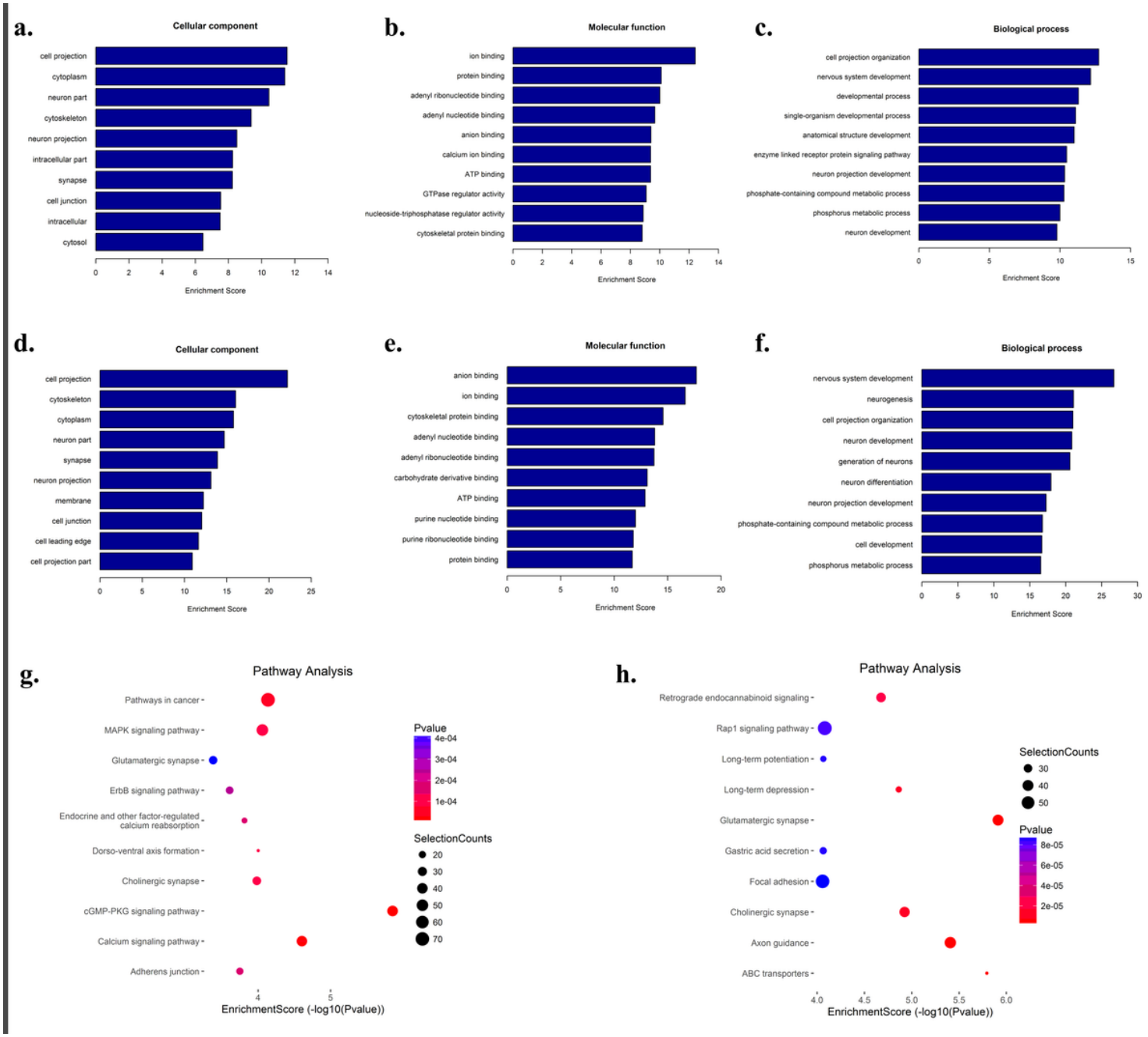

Figure 4

GO and KEGG pathway analysis based on the genes generating eccDNAs at differential level. a/b/c. The cellular components, molecular function, and biological processes based on the down-regulated eccDNAs. d/e/f. The cellular components, molecular function, and biological processes based on the up-regulated eccDNAs. g. KEGG pathway analysis based on the down-regulated eccDNAs in matched ESCC samples. h. KEGG pathway analysis based on the up-regulated eccDNAs in matched ESCC samples.

Figure 4

GO and KEGG pathway analysis based on the genes generating eccDNAs at differential level. a/b/c. The cellular components, molecular function, and biological processes based on the down-regulated eccDNAs. d/e/f. The cellular components, molecular function, and biological processes based on the up-regulated eccDNAs. g. KEGG pathway analysis based on the down-regulated eccDNAs in matched ESCC samples. h. KEGG pathway analysis based on the up-regulated eccDNAs in matched ESCC samples.

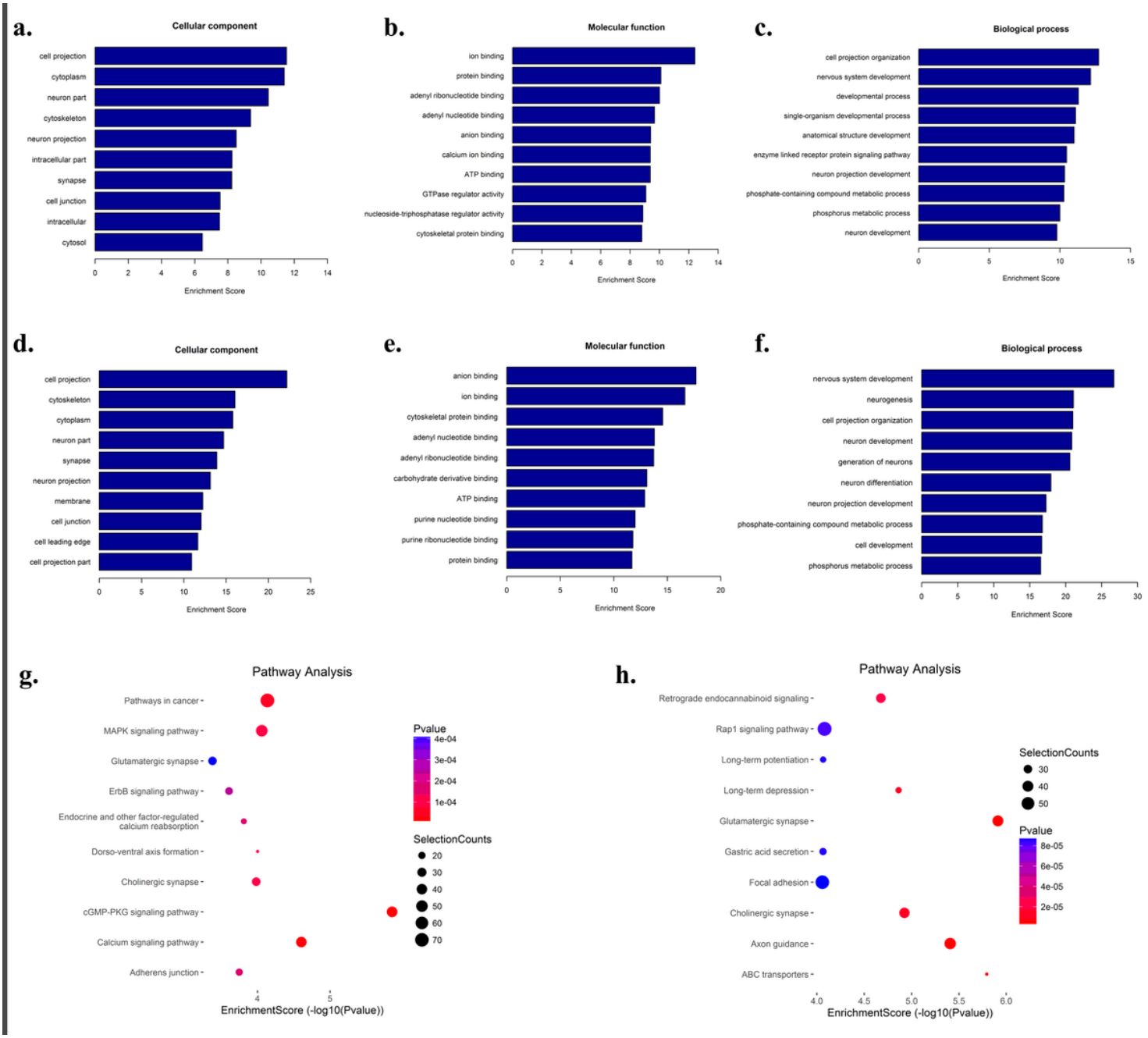

Figure 4

GO and KEGG pathway analysis based on the genes generating eccDNAs at differential level. a/b/c. The cellular components, molecular function, and biological processes based on the down-regulated eccDNAs. d/e/f. The cellular components, molecular function, and biological processes based on the up-regulated eccDNAs. g. KEGG pathway analysis based on the down-regulated eccDNAs in matched ESCC samples. h. KEGG pathway analysis based on the up-regulated eccDNAs in matched ESCC samples.

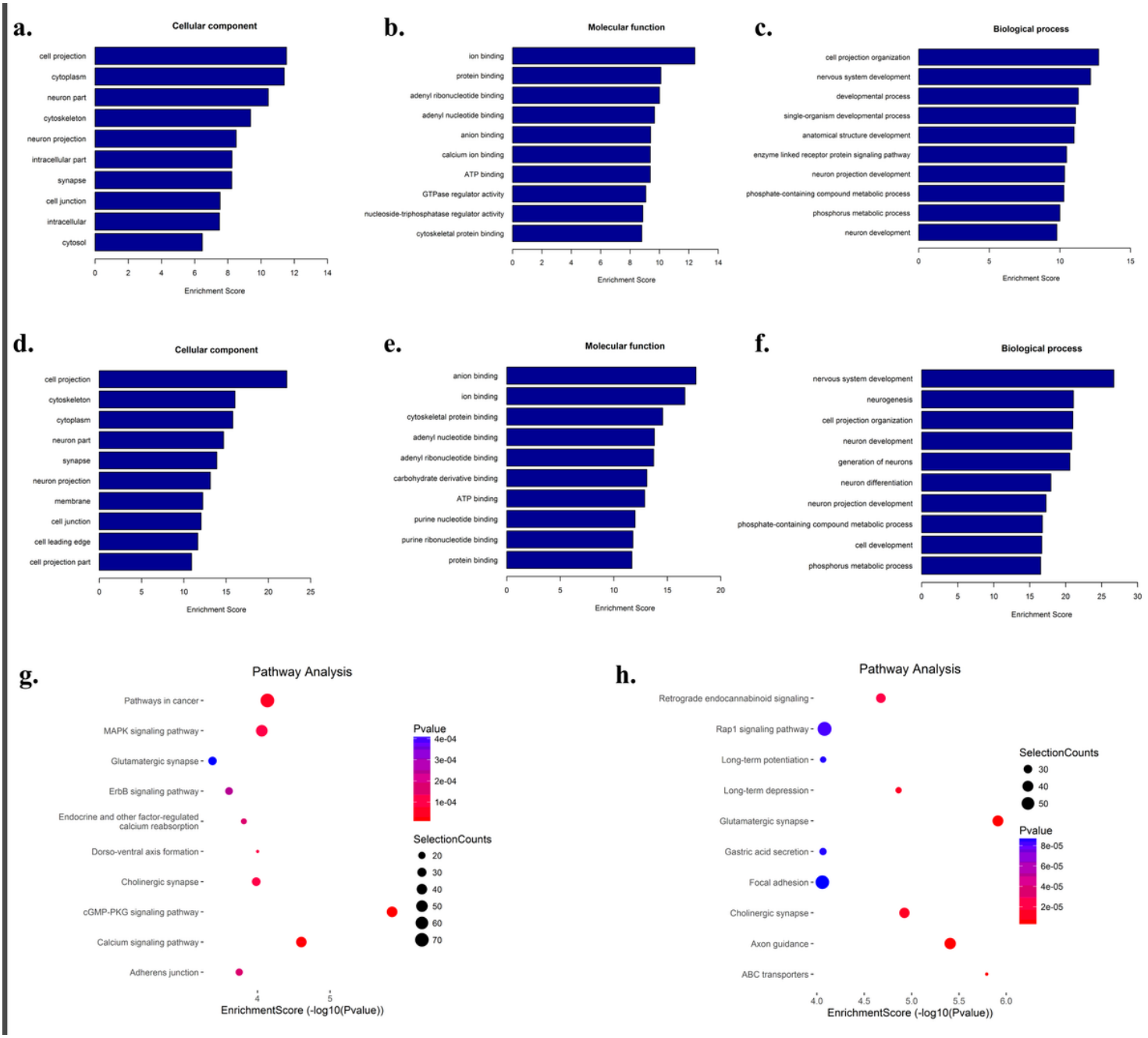

Figure 4

GO and KEGG pathway analysis based on the genes generating eccDNAs at differential level. a/b/c. The cellular components, molecular function, and biological processes based on the down-regulated eccDNAs. d/e/f. The cellular components, molecular function, and biological processes based on the up-regulated eccDNAs. g. KEGG pathway analysis based on the down-regulated eccDNAs in matched ESCC samples. h. KEGG pathway analysis based on the up-regulated eccDNAs in matched ESCC samples.

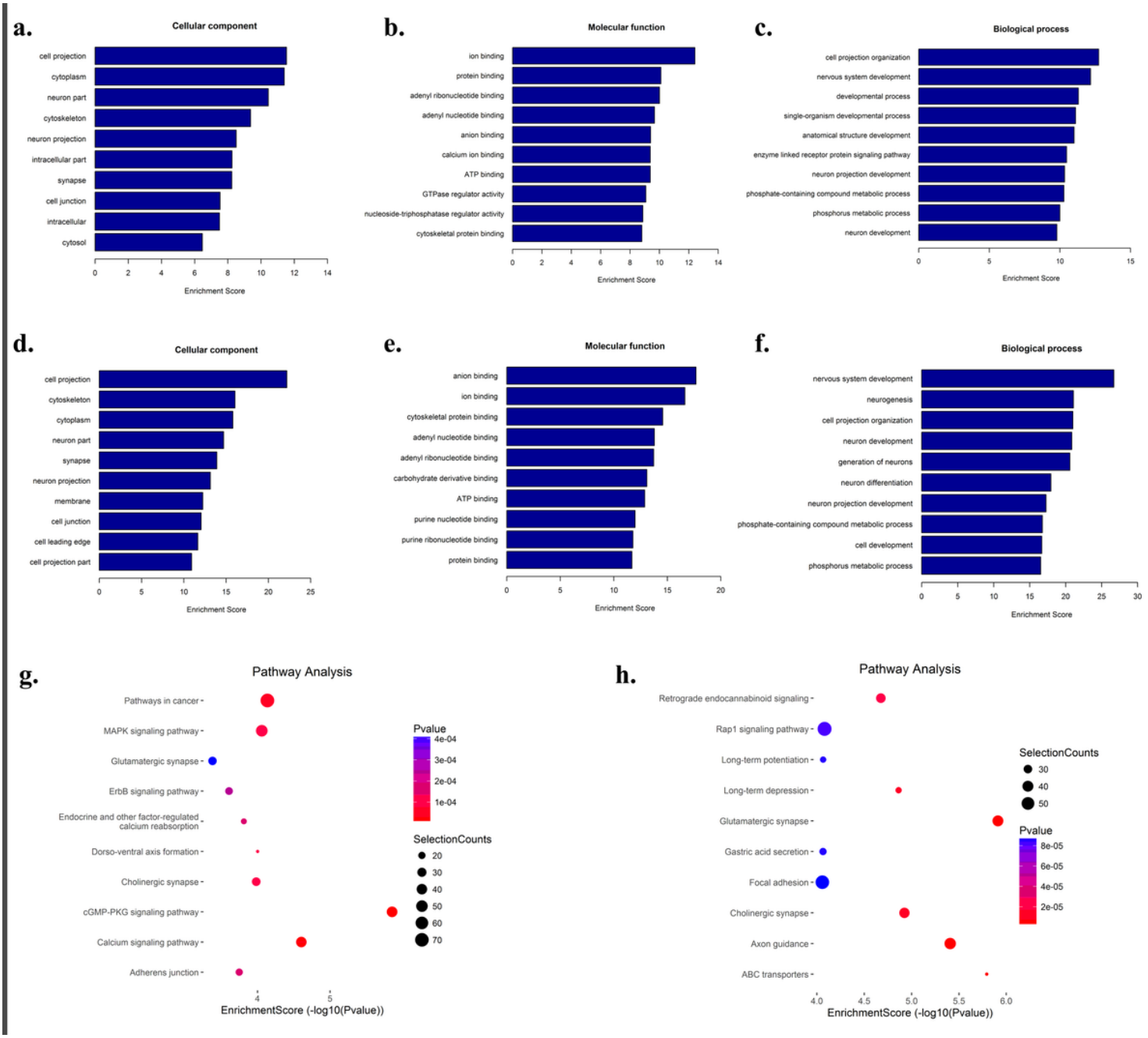

Figure 4

GO and KEGG pathway analysis based on the genes generating eccDNAs at differential level. a/b/c. The cellular components, molecular function, and biological processes based on the down-regulated eccDNAs. d/e/f. The cellular components, molecular function, and biological processes based on the up-regulated eccDNAs. g. KEGG pathway analysis based on the down-regulated eccDNAs in matched ESCC samples. h. KEGG pathway analysis based on the up-regulated eccDNAs in matched ESCC samples.

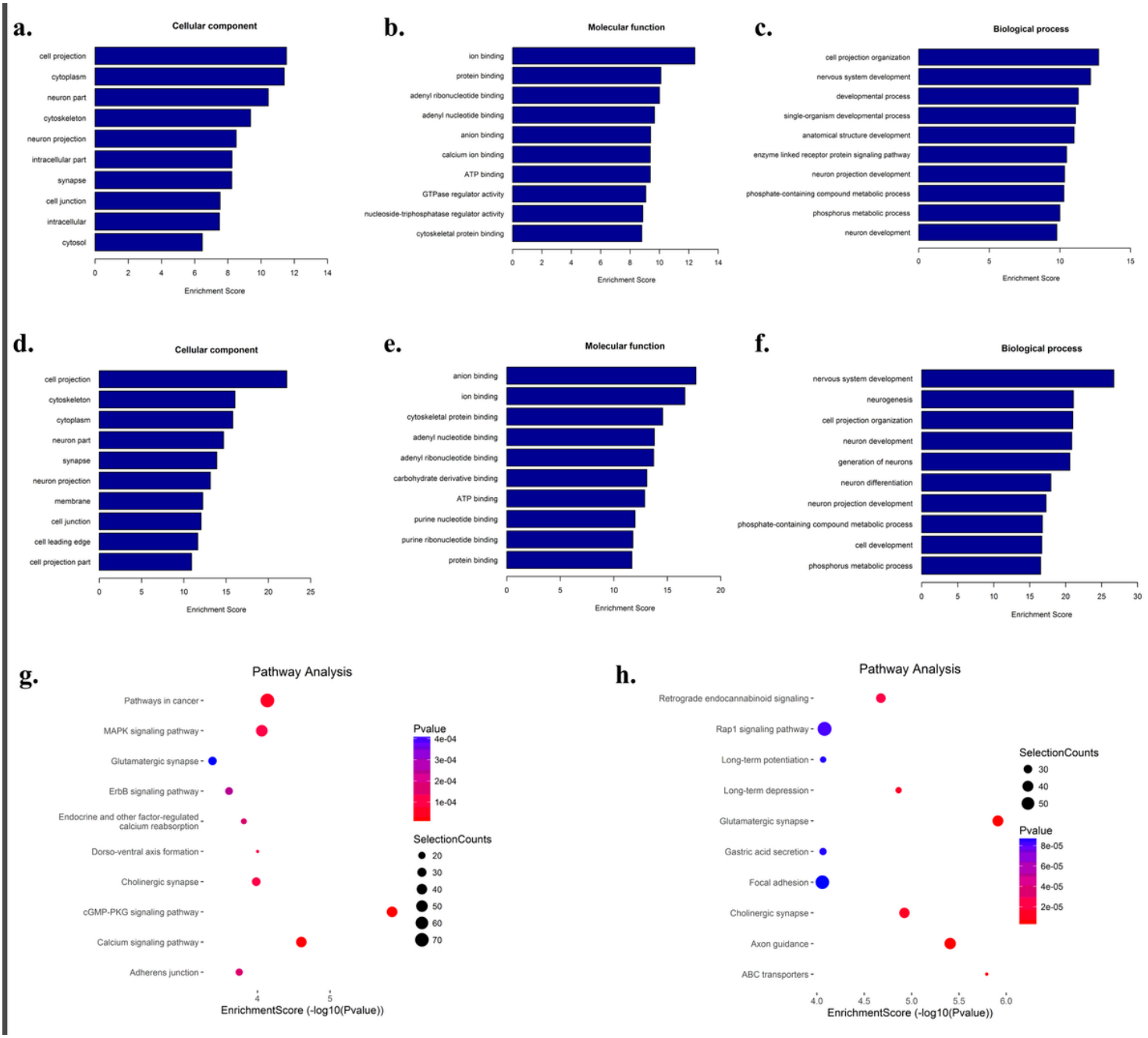

Figure 4

GO and KEGG pathway analysis based on the genes generating eccDNAs at differential level. a/b/c. The cellular components, molecular function, and biological processes based on the down-regulated eccDNAs. d/e/f. The cellular components, molecular function, and biological processes based on the up-regulated eccDNAs. g. KEGG pathway analysis based on the down-regulated eccDNAs in matched ESCC samples. h. KEGG pathway analysis based on the up-regulated eccDNAs in matched ESCC samples.

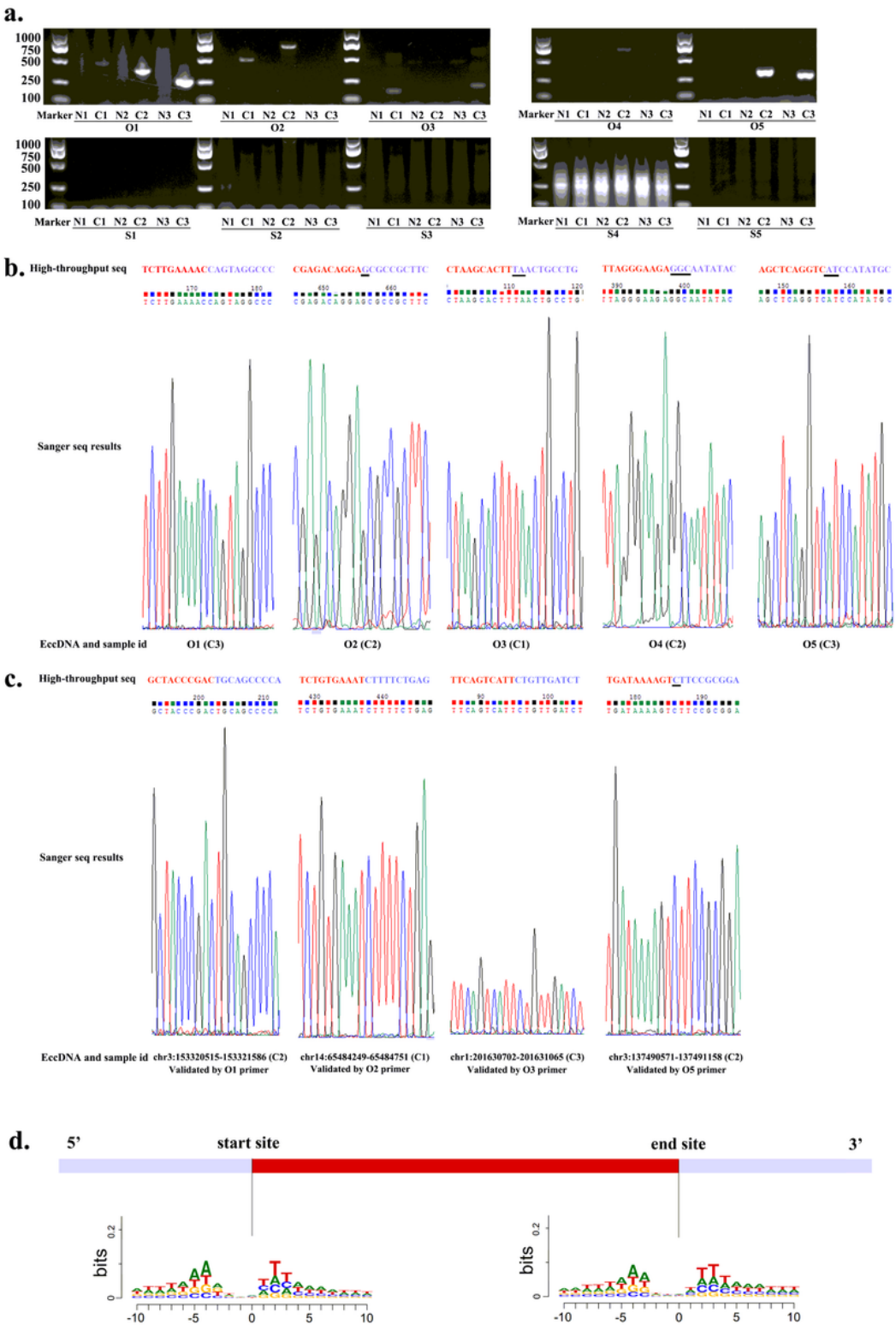

Figure 5

Validation of junctional sites of eccDNAs and motif analysis flanking junctional sites. a. The PCR results of 5 down-regulated and 5 up-regulated candidate eccDNAs in 3 pairs of matched ESCC samples. b. The results of Sanger sequencing validated the junctional site of the 5 up-regulated eccDNA in tumor samples. In high-throughput sequencing, red and blue color indicated the two ends of junctional sites, while the nucleotides with underscore indicated the same nucleotides at both the two ends. c. The results

of Sanger sequencing unexpectedly validated junctional sites of the other 4 true eccDNAs detected by high-throughput sequencing. d. The motif analysis flanking eccDNA junctional sites by high throughput sequencing.

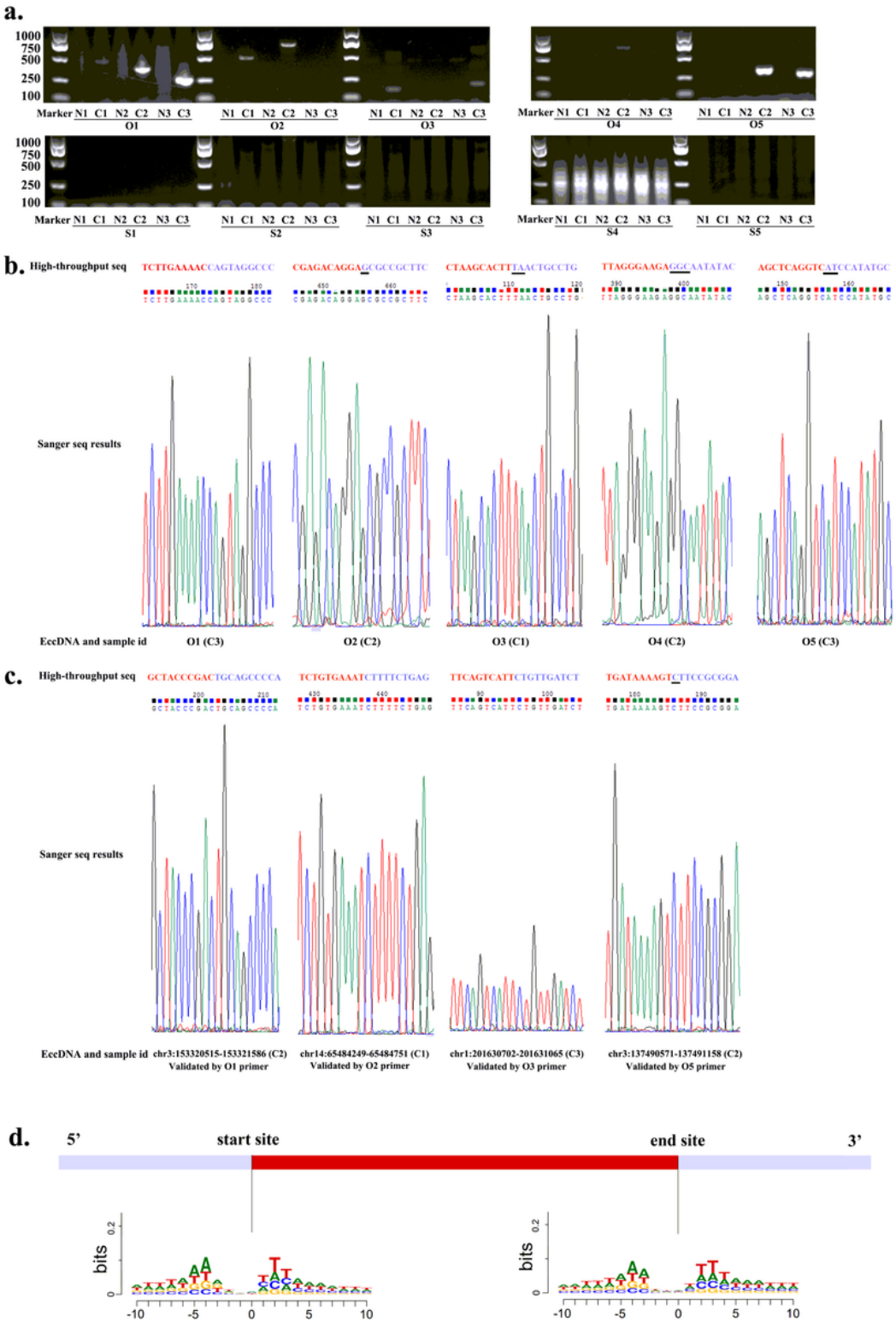

Figure 5

Validation of junctional sites of eccDNAs and motif analysis flanking junctional sites. a. The PCR results of 5 down-regulated and 5 up-regulated candidate eccDNAs in 3 pairs of matched ESCC samples. b. The

results of Sanger sequencing validated the junctional site of the 5 up-regulated eccDNA in tumor samples. In high-throughput sequencing, red and blue color indicated the two ends of junctional sites, while the nucleotides with underscore indicated the same nucleotides at both the two ends. c. The results of Sanger sequencing unexpectedly validated junctional sites of the other 4 true eccDNAs detected by high-throughput sequencing. d. The motif analysis flanking eccDNA junctional sites by high throughput sequencing.

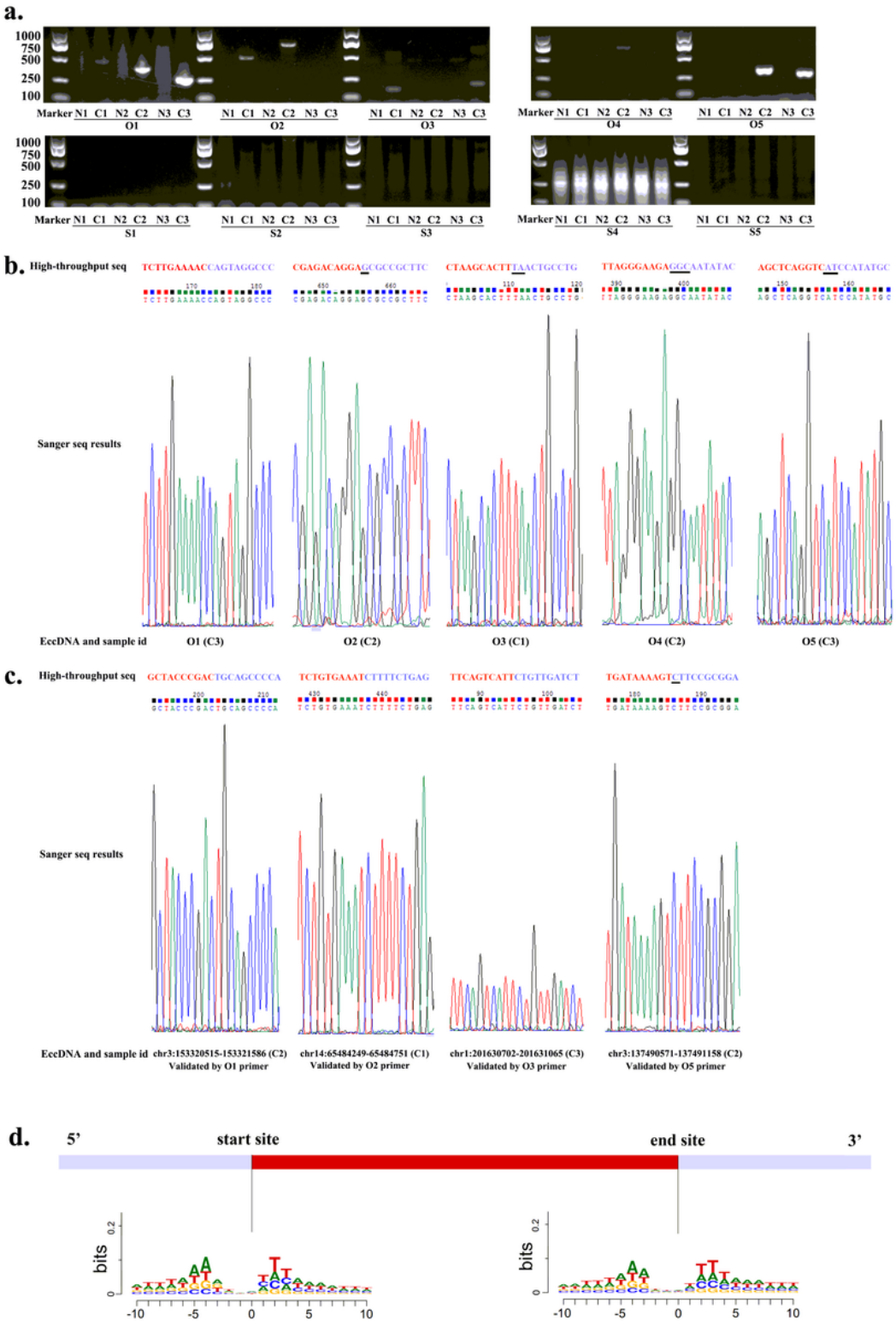

Figure 5

Validation of junctional sites of eccDNAs and motif analysis flanking junctional sites. a. The PCR results of 5 down-regulated and 5 up-regulated candidate eccDNAs in 3 pairs of matched ESCC samples. b. The results of Sanger sequencing validated the junctional site of the 5 up-regulated eccDNA in tumor samples. In high-throughput sequencing, red and blue color indicated the two ends of junctional sites, while the nucleotides with underscore indicated the same nucleotides at both the two ends. c. The results of Sanger sequencing unexpectedly validated junctional sites of the other 4 true eccDNAs detected by high-throughput sequencing. d. The motif analysis flanking eccDNA junctional sites by high throughput sequencing.

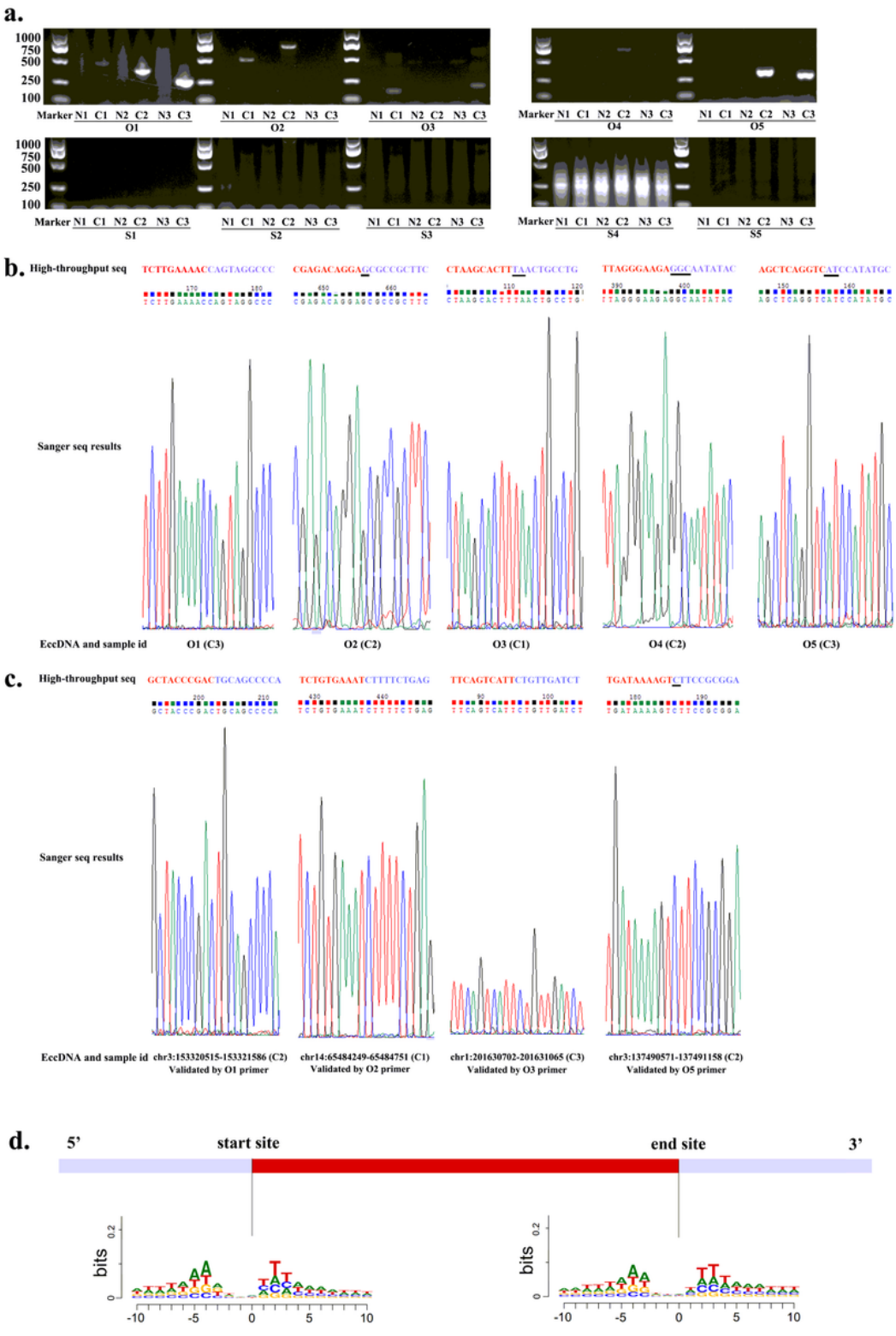

Figure 5

Validation of junctional sites of eccDNAs and motif analysis flanking junctional sites. a. The PCR results of 5 down-regulated and 5 up-regulated candidate eccDNAs in 3 pairs of matched ESCC samples. b. The results of Sanger sequencing validated the junctional site of the 5 up-regulated eccDNA in tumor samples. In high-throughput sequencing, red and blue color indicated the two ends of junctional sites, while the nucleotides with underscore indicated the same nucleotides at both the two ends. c. The results

of Sanger sequencing unexpectedly validated junctional sites of the other 4 true eccDNAs detected by high-throughput sequencing. d. The motif analysis flanking eccDNA junctional sites by high throughput sequencing.

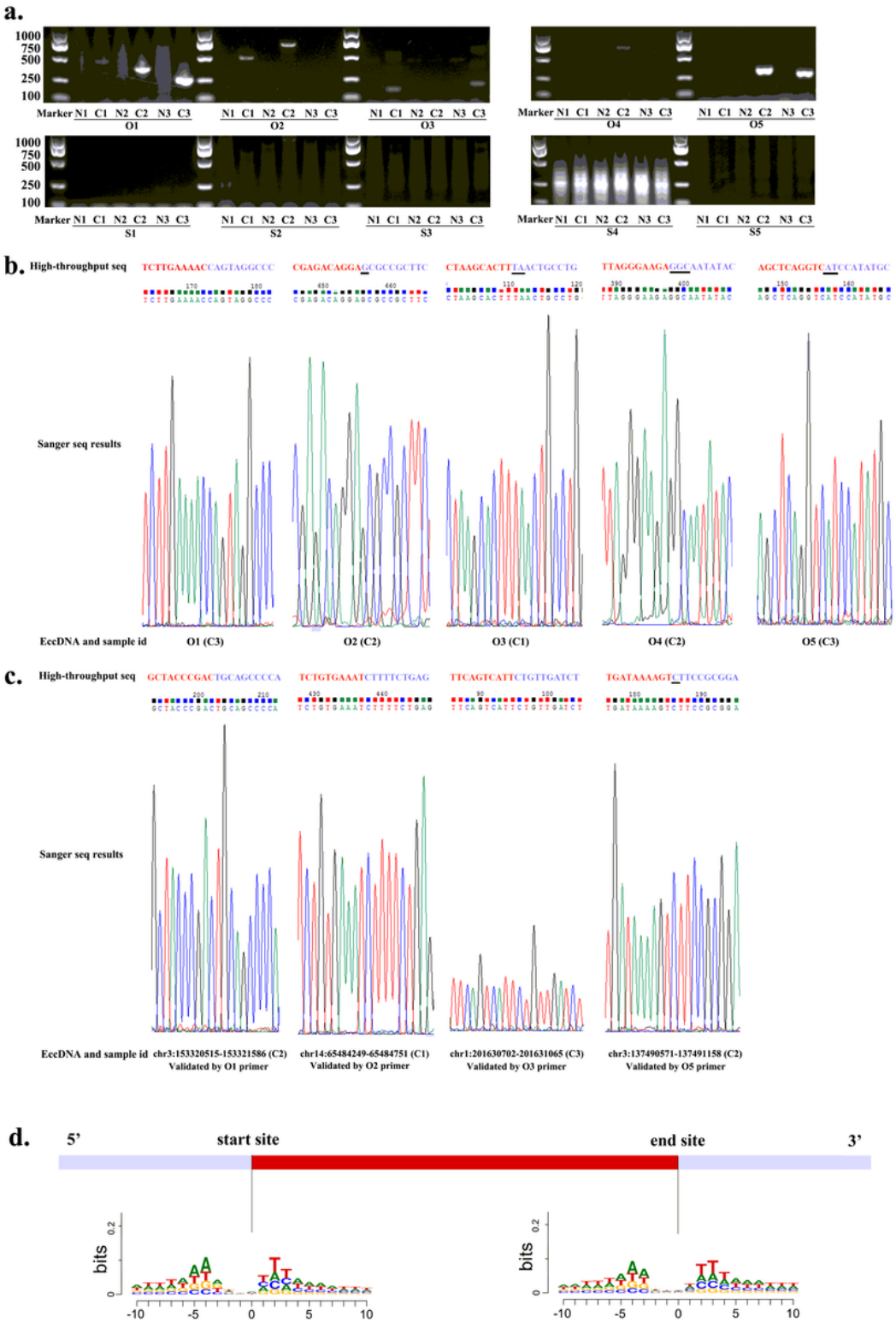

Figure 5

Validation of junctional sites of eccDNAs and motif analysis flanking junctional sites. a. The PCR results of 5 down-regulated and 5 up-regulated candidate eccDNAs in 3 pairs of matched ESCC samples. b. The

results of Sanger sequencing validated the junctional site of the 5 up-regulated eccDNA in tumor samples. In high-throughput sequencing, red and blue color indicated the two ends of junctional sites, while the nucleotides with underscore indicated the same nucleotides at both the two ends. c. The results of Sanger sequencing unexpectedly validated junctional sites of the other 4 true eccDNAs detected by high-throughput sequencing. d. The motif analysis flanking eccDNA junctional sites by high throughput sequencing.

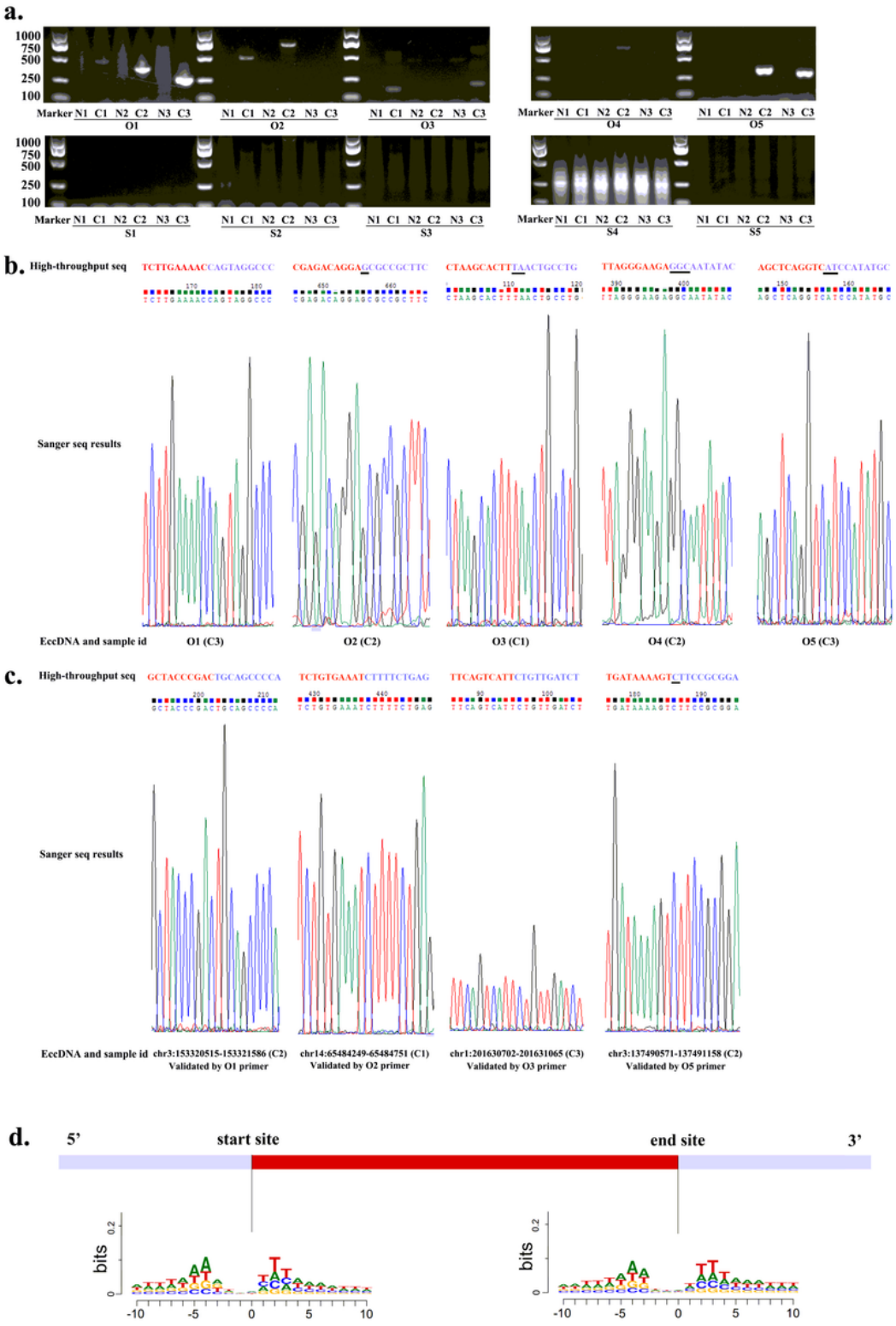

Figure 5

Validation of junctional sites of eccDNAs and motif analysis flanking junctional sites. a. The PCR results of 5 down-regulated and 5 up-regulated candidate eccDNAs in 3 pairs of matched ESCC samples. b. The results of Sanger sequencing validated the junctional site of the 5 up-regulated eccDNA in tumor samples. In high-throughput sequencing, red and blue color indicated the two ends of junctional sites, while the nucleotides with underscore indicated the same nucleotides at both the two ends. c. The results of Sanger sequencing unexpectedly validated junctional sites of the other 4 true eccDNAs detected by high-throughput sequencing. d. The motif analysis flanking eccDNA junctional sites by high throughput sequencing.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx
- SupplementaryMaterials.docx
- SupplementaryMaterials.docx
- SupplementaryMaterials.docx
- SupplementaryMaterials.docx
- SupplementaryMaterials.docx