Cellular and transcriptional heterogeneity in the intrahepatic biliary epithelium

Hannah R Hrncir¹,², Adam D Gracz¹,²,³

Affiliations:
³ Department of Medicine, Division of Digestive Diseases
² Graduate Program in Biochemistry, Cell and Developmental Biology
³ Graduate Program in Genetics and Molecular Biology

Emory University, Atlanta, GA

Correspondence: agracz@emory.edu

Authors’ contributions: HRH generated figures and tables. HRH and ADG drafted and edited the manuscript.

Conflicts of Interest: The authors disclose no conflicts.

Funding: None

Ethical Statement: The study did not require the approval of an institutional review board
Abstract

Epithelial tissues comprise heterogeneous cellular subpopulations, which often compartmentalize specialized functions like absorption and secretion to distinct cell types. In the liver, hepatocytes and biliary epithelial cells (BECs; also called cholangiocytes) are the two major epithelial lineages and play distinct roles in (1) metabolism, protein synthesis, detoxification, and (2) bile transport and modification, respectively. Recent technological advances, including single cell transcriptomic assays, have shed new light on well-established heterogeneity among hepatocytes, endothelial cells, and immune cells in the liver. However, a “ground truth” understanding of molecular heterogeneity in BECs has remained elusive, and the field currently lacks a set of consensus biomarkers for identifying BEC subpopulations. Here, we review long-standing definitions of BEC heterogeneity as well as emerging studies that aim to characterize BEC subpopulations using next generation single cell assays. Understanding cellular heterogeneity in the intrahepatic bile ducts holds promise for expanding our foundational mechanistic knowledge of BECs during homeostasis and disease.

Introduction

Cellular heterogeneity supports tissue function in homeostasis and provides tissues with the ability to adapt to environmental challenges. The gastrointestinal (GI) tract holds many well-defined examples of cell type specific functions that aid in digestion and homeostasis. Well-characterized intestinal epithelial cells (IECs) demonstrate specialized functions spanning absorption of nutrients to secretion of hormones, mucus, or antimicrobial peptides [1]. Epithelial heterogeneity also varies regionally along the small intestine and colon and appears to be at least partially cell-autonomous, with cells maintaining their identity when transplanted from the small intestine to the colon [2]. Pancreatic beta cells provide another example of GI epithelial heterogeneity, with subpopulations responding to different thresholds of glucose, exhibiting varied gap junction expression patterns, and demonstrating differential proliferative capacity [3]. In both IECs and pancreatic beta cells, recent single cell transcriptomic studies have identified previously unappreciated heterogeneity, generating exciting new insight on GI epithelial function [4-6]. Refining models of cellular heterogeneity and regulation of cellular identity can provide foundational mechanistic understanding of tissue function and may help identify cellular and molecular targets for novel therapeutics.

Hepatocytes and BECs represent the two epithelial lineages of the liver. Hepatocytes make up ~80% of the liver by weight and are responsible for a variety of functions including glucose metabolism, detoxifying exogenous compounds, and bile salt secretion [7]. BECs are polarized epithelial cells that line the biliary tree and are responsible for regulating net bile volume, modifying bile, and maintaining bile alkalinity [8]. BECs represent only 3-5% of the liver by mass but can produce up to 40% of daily bile output volume [9]. While normally quiescent, BECs demonstrate extensive remodeling and proliferation following liver injury, in a process called “ductular reaction” [10-12]. Though BECs of the intrahepatic and extrahepatic biliary tree share many biomarkers and make up a continuous ductal network, these cells have distinct embryological origins [13, 14]. In this review, we focus on BEC heterogeneity in the intrahepatic bile ducts (IHBDs).

Hepatocyte heterogeneity during homeostasis is well-defined in relation to hepatocyte zonation, which refers to the functional compartmentalization of specific metabolic processes in the liver that is regulated through signaling gradients [15-17]. In contrast, classical definitions of BEC heterogeneity have
focused on morphological characteristics rather than molecular heterogeneity. The recent use of novel biomarkers and emerging single cell technologies has characterized the molecular identity of individual BECs and shed new light on cellular subpopulations of the IHBDs. These studies support the existence of multiple BEC subpopulations with distinct morphological, functional, and molecular properties.

**Bile duct anatomy**

BECs line the bile ducts that, collectively, form the biliary tree. The biliary tree is subdivided into extrahepatic bile ducts (EHBDs) and IHBDs. In humans, EHBDs are composed of the common hepatic duct, cystic duct, common bile duct, main pancreatic duct, and the ampulla of Vater. Human IHBDs consist of hepatic ducts (>800 μm diameter), segmental ducts (400–800 μm diameter), area ducts (300–400 μm diameter), septal ducts (100 μm diameter), interlobular ducts (15–100 μm diameter), and bile ductules (<15 μm diameter) [18-20]. These ducts can be grouped by duct diameter into small and large ducts. Small ducts consist of septal ducts, interlobular ducts, and bile ductules, while large ducts consist of area, segmental, and hepatic ducts. Though they lack the exact structural complexity observed in human IHBDs, rodent IHBDs mimic this variation in size with large (rat: >15 μm diameter) and small ducts (rat: <15 μm diameter)[21, 25]. Human and rat EHBDs and IHBDs also contain peribiliary glands (PBGs), which are small invaginations along the bile ducts that secrete mucus and other compounds [64]. PBGs are thought to serve as niches for immature BECs, as they exhibit a hyperproliferative response to injury and can give rise to biliary tract cancers [34, 65-67]. While mice have well-characterized PBGs along the EHBD, PBGs have not been characterized in mouse IHBDs [26, 27].

IHBDs play an important role in modifying primary canalicular bile produced by hepatocytes and funneling bile into the EHBDs. BECs perform secretory and absorptive functions, which adjust bile flow and alkalinity to meet physiological needs. When bile removal is seriously altered in the context of cholestatic liver injury or disease, progressive cirrhosis can occur and lead to liver failure. Cholangiopathies, liver diseases which directly impact BECs, account for additional morbidity and mortality given their progressive nature and lack of effective medical therapies [28]. Some cholangiopathies can affect the bile ducts heterogeneously. For example, primary sclerosing cholangitis (PSC) varies in etiology with most cases affecting large ducts and a subset of cases affecting small ducts (~10%) [29]. Small duct PSC is associated with better long-term outcomes compared to large duct PSC [30-32]. In a minority of PSC cases, both small and large ducts are affected [30]. In contrast, primary biliary cholangitis (PBC) is an autoimmune disorder resulting in progressive IHBD destruction specific to small ducts [33, 34]. The heterogeneous nature of PBC and PSC demonstrates the need for an improved understanding of IHBD heterogeneity and how this heterogeneity promotes proper duct function.

**Comparative anatomy of model organisms for studying intrahepatic bile ducts**

Standard animal models for studying IHBDs are zebrafish, rats, and mice. The existence of distinct liver lobes, referred to as lobation, is a common structural variation between species (Table 1). Zebrafish have bilobated livers (dorsal and ventral lobes) while mouse and rat livers consist of median, left, right, and caudate lobes (Table 1) [35, 36]. Generally, liver size and lobation is influenced by organism size with smaller mammals presenting lobated livers (i.e. mice, rats, and cats) and larger mammals possessing non-lobated livers (i.e. humans, dolphins, and cattle) [37]. In mammals, the hepatic artery, portal veins, and bile ducts are organized into a portal triad, with hepatocyte plates (also known as chords) extending away from the triad. Mammals, including humans, rats, and mice, have conserved IHBD development.
Hepatic progenitor cells, called hepatoblasts, give rise to hepatocytes and BECs through asymmetric specification [38]. During biliary tubulogenesis, early BECs align around the portal vein to form a single layered ring termed the ductal plate. Portions of the ductal plate become a bilayer and form expanding luminal structures which give rise to bile ducts [39-42]. BEC specification from hepatoblasts is promoted by several pathways, including Notch, TGF-β, and Hippo, which may also play roles in regulating BEC heterogeneity [42-45].

Although rat liver development is very similar to mouse and human liver development, rats do not develop a gallbladder or cystic duct, due likely to a lack of Sox17 in the biliary bud (Table 1) [46]. In teleosts, including zebrafish, there is no evidence of a ductal plate [47]. Instead, BECs first appear near the extrahepatic bile duct and expand to the rest of the liver [48]. Resulting from these developmental differences, zebrafish lack portal triads and instead have bile ducts dispersed throughout the liver lobule (Table 1). These luminal structures gradually enlarge in size, eventually joining larger ducts which carry bile to the common bile duct [49]. To date, there is no evidence of a hepatic artery in zebrafish, though some studies have described small hepatic artery radicles [49-51]. Zebrafish BEC heterogeneity remains the least well-characterized among common animal models of liver biology but has been studied morphologically (small and large ducts) and recently, through scRNA-seq (Table 1) [52, 53].

Although zebrafish have substantial structural differences in duct organization relative to mammals, they present some technical advantages for modeling IHBD development. Since zebrafish embryos are transparent, ductal development can be observed easily using whole mount brightfield microscopy or fluorescent microscopy to visualize transgenic reporters [51]. Zebrafish also develop quickly, with the first BECs appearing at 50 – 60 hours post-fertilization (hpf) and duct-like structures forming 70 hpf [49]. Critical BEC developmental gene expression such as NOTCH activation and the expression of Hnf6, Sox9 (sox9b in zebrafish), and Onecut3 is conserved in all vertebrates including zebrafish [49, 54]. Transparency, quick gestation time, high fecundity, and conserved developmental genetics allow zebrafish to be a useful tool in unraveling the importance and molecular regulation of BEC heterogeneity in addition to studying liver development and liver disease mechanisms.

Although zebrafish are a useful model, mice and rats are advantageous in that rodent liver development and organization is largely similar to humans. However, due to expression of Cyp2c70 which converts chenodeoxycholic acid (CDCA) to 6-hydroxylated muricholic acids (MCAs) by 6-hydroxylation, rodent bile is more hydrophilic than human bile [55, 56]. The differences in bile acid composition between rodents and humans may be important to consider when comparing differences in human and rodent BEC heterogeneity, especially if aspects of cell-cell heterogeneity are impacted by bile acid signaling. In rodents, many well-established injury models and transgenic mouse lines facilitate assays of liver homeostasis and repair and simplify genetic manipulation [57-59]. Techniques for producing organoids from ductal fragments or single BECs have been well characterized in mice, rats, and patient samples [60, 61]. Ductal organoids are complementary systems to in vivo studies of bile duct development, disease, and heterogeneity and are amenable to genetic manipulation and drug screening. To study BEC heterogeneity, single cell organoids can be grown in culture to determine differences in proliferation, survival, and responses to exogenous factors introduced into the culture system [62]. While BEC organoids are still evolving as a model system, they are likely to prove useful in unraveling mechanisms regulating BEC heterogeneity.

Classical definitions of BEC morphological heterogeneity
Intrahepatic BEC heterogeneity was first described in the context of duct size [21-24]. BECs that line small ducts were observed to be significantly smaller and less columnar than BECs that line large ducts (Table 2). This led to the classification of BECs as “small” or “large”, corresponding to the relative diameter of the resident duct as well as the area of the BECs themselves [63, 64]. Small BECs have also been characterized as having a high nucleus to cytoplasm ratio, while large BECs exhibit a more extensive cytoplasm [64].

Primary cillum length is another structural difference between small and large BECs. The apical cell membrane of BECs possess a single primary cillum which has sensory function and regulates biological activities such as cell differentiation, proliferation, and secretion [65, 66]. Small BECs have been reported to have shorter cilia, which are less than half the size of cilia found on large BECs (Table 2; in rats: 3.58 ± 1.12 μm and 7.35 ± 1.32 μm, respectively) [66, 67]. However, there is not a clear consensus in the literature regarding the percentage of BECs which possess a primary cillum. Some reports describe as few as 43.5 ± 2.0% of BECs possessing primary cilia in rats [68]. Another study has detected a similar heterogeneity in mouse BEC cilia, reporting ~ 0.3 cilia per duct cell [69]. Other studies demonstrate 80 - 100% of BECs possessing primary cilia in mice and rats [70, 71]. Differences in reports of cilia count may be due in part to histological artifacts, as some cilia could be excluded based on the orientation and thickness of the duct captured in the tissue section. Further examination of primary cillum presence and length may be warranted in future studies characterizing structural heterogeneity among BECs.

Functional heterogeneity in small and large BECs

In addition to distinct morphological characteristics, small and large BECs also exhibit functional differences in terms of how they modify bile. General BEC secretory and absorptive function has been extensively reviewed elsewhere; here we focus on functional differences between small and large BECs [72]. In IHBDs, intracellular calcium (Ca^{2+}) modulates diverse processes including cell proliferation, apoptosis, bile alkalization, and secretion of electrolytes and water [73-75]. In BECs, Ca^{2+} release is regulated predominately by 1,4,5-trisphosphate (InsP3), which binds to InsP3 receptors (InsP3R) to promote Ca^{2+} release from the ER [73, 74]. While both small and large BECs can signal through InsP3Rs, they differ in Ca^{2+} mobilizing receptors. In small ducts, histamine receptor H1 (HRH1) is found in the basolateral membrane and modulates small BEC proliferation and secretion (Figure 1) [76, 77]. Small BECs also regulate proliferation by Ca^{2+}-dependent activation of the transcription factors: NFAT2, NFAT4, and SPI (Figure 1)[78]. Similarly, while they have been reported to lack HRH1, large BECs express HRH2, which has been shown to be a key regulator of large BEC proliferation (Figure 1) [76, 77, 79, 80]. Large BECs also have Ca^{2+}-dependent activation of the transcription factor NFAT3 [78].

Unlike small BECs, large BECs have been reported to be cAMP responsive and regulators of bile flow. Bile flow is controlled by large BEC specific receptors: cystic fibrosis transmembrane conductance regulator (CFTR), Cl^+/HCO_3^- anion exchanger 2 (AE2), secretin receptor (SCTR), somatostatin receptor (SSTR), melatonin receptors (MT1 and MT2), and apical sodium-dependent bile acid transporter (ASBT). Secretin regulates large BEC bicarbonate (HCO_3^-) secretion through interaction with the G protein coupled SCTR, resulting in cAMP signaling [24, 81]. cAMP signaling induces CFTR, which activates AE2 and Ca^{2+} is released into bile (Figure 1) [23, 82, 83]. Somatostatin acts to counter-regulate the stimulatory effects of secretin by binding SSTR leading to inhibition of ductal fluid secretion, and in some instances,
ductal fluid absorption [79, 84-88]. cAMP signaling is also decreased in large BECs by melatonin, which has been shown to suppress BEC proliferation during cholestasis. Melatonin is found in bile (2-3 times higher than in circulation and other bodily fluids) and interacts with large BEC specific melatonin receptors (MT1 and MT2) (Figure 1) [89, 90]. Bile acid (BA) is transported inside BECs by ASBT [91]. Once inside the cell, BAs bind the ileal bile acid binding protein (tIBABP). ASBT and IBABP have been reported to be expressed at higher levels in large BECs but are present in small BECs [91]. The functional differences between small and large BECs suggest that large ducts may be more important than small ducts in regulating bile flow due to large BEC responsiveness to secretin, somatostatin, melatonin, and BAs.

In addition to heterogeneous mechanisms for mobilizing Ca$^{2+}$, BECs exhibit functional differences in proliferation rates following injury. While BECs remain largely quiescent in the normal liver and homeostatic turnover is poorly understood, proliferative heterogeneity of BECs has been studied during injury and regeneration [92]. Animal models of ductular reaction rely on a range of established surgical and chemical damage models, often resulting in cholestasis secondary to hepatocyte injury (common damage models are reviewed extensively in [93]). In both mice and rats, bile duct ligation (BDL) reportedly stimulates proliferation of large - but not small - BECs. These data were obtained by measuring duct size in tissue sections and measuring relative cell size and cell cycle using flow cytometry [21, 86]. In contrast, IHBD proliferation following carbon tetrachloride (CCl$_4$) feeding in rats occurs primarily in small BECs, as assayed by measuring the number of BECs which were positive for proliferating cell nuclear antigen (PCNA) in tissue sections [94]. Another study found a significant increase in both small and large BEC proliferation following bile acid feeding in mice, suggesting that this model does not induce a heterogenous proliferative response between small and large ducts [95].

Proliferative heterogeneity in BECs has also been assayed using an in vivo clonal labeling technique. Kamimoto et al., used Pro1$^{CreERT2}$,R26$^{tdTomato}$ transgenic mice to track proliferation via inducible, BEC-specific recombination and tdTomato expression [96]. By administering a low dose of tamoxifen, BECs were sparsely labelled, allowing the proliferation and clonal expansion of individual BECs to be traced over time. BECs were labeled at a rate of 0.2% and following 6 weeks of thioacetamide (TAA) injury, most BECs exhibited no or minimal proliferation. However, rare BEC clones had formed colonies of dozens of cells at the same timepoint [96]. These findings clearly demonstrate that BECs do not proliferate uniformly following damage. The authors also sought to determine if BEC proliferation is anatomically compartmentalized. Rather than using established “small” and “large” duct definitions, the authors defined ductules as < 8μm in diameter and ducts as > 8μm, based on anatomical relationships between the IHBDs and portal veins observed by 3D microscopy [96]. The authors reported no large clones of proliferating BECs in the ducts. Instead, large proliferative clones were localized to peripheral ductules [96]. These results differ from earlier findings that reported proliferation following BDL to be restricted to large BECs and instead bear similarity to small BEC proliferation following CCl$_4$ injury in rats. Collectively, these studies suggest that proliferative heterogeneity among BECs is likely influenced by the nature of the damage modality inducing ductular reaction. New and ongoing ductular reaction studies should consider small and large duct heterogeneity so that comparisons can be made to long standing definitions of biliary heterogeneity.

Genetic biomarkers for studying BEC heterogeneity
Cell type-specific biomarkers have long been advantageous for identifying epithelial subpopulations in other GI tissues and are often related to specialized cellular function. Well-accepted examples include lysozyme expression specific to small intestinal Paneth cells and glutamine synthetase expression specific to pericentral hepatocytes [97, 98]. While many studies on small and large BECs have focused on differential expression of specific genes and proteins, these markers are rarely used to identify or describe BEC subpopulations. A potential challenge is that many classic BEC physiology studies utilized rat tissue or immortalized cell lines, which may not directly translate to mouse models used in more recent genetic studies. Additionally, many of the most well-accepted BEC markers, including EPCAM, K19, and SOX9, appear to be expressed ubiquitously throughout IHBDs and are therefore unsuitable for pursuing questions of BEC heterogeneity. Flow cytometry/FACS and three-dimensional immunofluorescence demonstrate that CD326 reliably labels all cells of the IHBD in mice [99-104]. Similarly, SOX9 has been proposed as a marker of immature BECs but is an early transcriptional regulator of BEC specification during development and is expressed in all adult BECs of the IHBDs [42, 102, 105-107]. SOX9 is also expressed in rare peribiliary hepatocytes (termed “hybrid hepatocytes” or HybHeps) and may promote hepatocyte-to-BEC transdifferentiation during injury and tumorigenesis [102, 107-109].

ST14 was identified as a marker of BEC heterogeneity through a targeted screen of potential cell surface markers for cells expressing MIC1-1C3, an oval cell antigen that labels BECs [110]. MIC1-1C3+ cells were found to contain ST14lo and ST14hi subpopulations. A clonogenic biliary organoid assay demonstrated that while ST14lo cells could form organoids at P0, only ST14hi cells are capable of efficiently generating organoids that survive for greater than two passages [111]. Further, RNA-seq of ST14lo vs. ST14hi cells identified 899 differentially expressed genes. One such candidate marker of ST14hi cells, Pkhd11I, was examined by single molecule FISH (smFISH) and determined to be expressed in a minority of IHBD BECs. While the study reports that Pkhd11I+ cells appear dispersed throughout the IHBDs, their localization to small or large ducts was not reported [111].

With the advent of transgenic mice, knowledge of biomarkers was leveraged with fluorescent reporters to identify and isolate specific cell populations. The use of Lgr5EGFP mice to identify GI epithelial stem cell populations, first in the small intestine and colon and later in the stomach, is perhaps the most broadly applied example of this strategy [112, 113]. Of note, while Lgr5 was initially postulated to be expressed in a subpopulation of BECs based on Lgr5EGFP expression and organoid-forming capacity, subsequent studies have failed to identify Lgr5 mRNA in BECs [45, 114-116]. Rather, Lgr5 and Lgr5EGFP appear restricted to hepatocytes, and whether or not Lgr5+ hepatocytes have a privileged role in liver growth or regeneration remains controversial [117, 118].

Our lab recently applied a Sox9EGFP transgenic mouse model to explore BEC heterogeneity. The Sox9EGFP allele was developed as a part of the GENSAT Project and previously used to isolate distinct intestinal epithelial subpopulations, including stem and progenitor cells [119-122]. Biliary epithelial expression of Sox9 and the EGFP reporter is broad, with all EPCAM+ or K19+ BECs expressing EGFP [102]. However, distinct expression levels of the EGFP transgene facilitate identification and isolation of BEC subpopulations. Specifically, EGFPlo and EGFPhi cells represent transcriptomically-distinct BEC subpopulations that also demonstrate differential functional characteristics in clonal organoid-forming assays. While EGFPlo and EGFPhi BECs were found in both small and large IHBDs, smaller ductules were more enriched for EGFPhi BECs. EGFPhi BECs were also significantly smaller in area relative to EGFPlo BECs, further suggesting consistency between this population and small BECs. As expected from the known expression of Sox9 in HybHeps, cells expressing a very low level of the EGFP reporter (termed
EGFP\textsuperscript{sub}) were demonstrated to be consistent with peribiliary hepatocytes [102]. All EGFP\textsuperscript{+} populations, including EGFP\textsuperscript{sub} HybHeps, were capable of forming organoids in biliary and hepatocyte organoid conditions. Interestingly, organoids from EGFP\textsuperscript{sub} and EGFP\textsuperscript{hi} populations were more likely to express HNF4A in hepatocyte media conditions [102, 123]. Though hepatocyte function was not assayed, this may suggest differences in transdifferentiation potential between Sox9\textsuperscript{EGFP} populations. While the Sox9\textsuperscript{EGFP} allele may aid in dissecting BEC heterogeneity, it is limited in that it requires the presence of a transgenic allele and may not capture the full spectrum of BEC subpopulations.

**Transcriptomic heterogeneity and emerging single cell technologies in BEC biology**

The emergence of transcriptomic technologies, beginning with microarrays in the 1990s, has long supported biomarker discovery and studies of cellular heterogeneity [124]. However, determining transcriptomic heterogeneity between cell types through “bulk” assays, which rely on pooled samples of many individual cells, presents a circular problem. Bulk assays rely on well-defined biomarkers or techniques to fraction cells into known, distinct populations. Early attempts to characterize transcriptome wide BEC gene expression applied microarray assays to immortalized small and large BEC lines from mouse liver. These studies demonstrated substantial transcriptomic differences between the cell types, including confirmatory differential expression of genes previously reported to distinguish small and large BECs [125]. However, like all studies relying on immortalized cell lines, it is unclear to what degree in vitro cultivation impacted transcriptomic results. More recently, BEC subpopulations isolated by ST14 or Sox9\textsuperscript{EGFP} expression were subjected to RNA-seq, once again demonstrating substantial transcriptomic differences between BEC populations presumed to be distinct [102, 111]. While these studies are important in that they provide conceptual proof-of-concept that all intrahepatic BECs are not transcriptomically homogeneous, they all failed to identify or validate distinct genetic biomarkers of BEC subpopulations that have been adopted widely in the field.

While biomarkers are an important tool for identifying and studying cellular heterogeneity, they remain intrinsically limited by the fact that genes are rarely, if ever, expressed in a single cell type. In recent years, rapid technological advancements in single cell RNA-seq (scRNA-seq) have facilitated unprecedented studies of single cell transcriptomes in increasing numbers of cells [126-128]. Similarly rapid gains in computational tools have allowed cell biologists to identify and validate novel cellular subpopulations in scRNA-seq without any a priori knowledge of cell type specific biomarkers. Such studies have led to remarkable findings, including the identification of many new cellular subpopulations within cell types that were previously thought to be relatively homogenous, including intestinal tuft cells and liver sinusoidal endothelial cells [5, 129]. Due to its central relevance to organismal homeostasis and involvement in a wide range of human pathological conditions, the liver has been the focus of several large scRNA-seq studies aimed at generating cell “atlases” [130-133]. However, because liver cell isolation protocols do not often provide robust isolation of BECs, they are frequently underrepresented in “whole liver” scRNA-seq datasets. Regardless, the first human liver cell atlas project examined EPCAM\textsuperscript{+} cells in an attempt to identify putative bipotent hepatic progenitors [130]. An analysis of 1,087 EPCAM\textsuperscript{+} cells demonstrated the existence of subpopulations enriched for conventional hepatocyte or BEC genes. Further, TROP2 was identified as a putative progenitor marker, with organoid-forming capacity restricted to EPCAM\textsuperscript{+} cells co-expressing intermediate or high levels of TROP2 [130]. This led to the conclusion that EPCAM\textsuperscript{+}/TROP2\textsuperscript{int} cells represent bipotent human liver progenitors, which may require additional characterization in light of emerging evidence that liver regeneration is driven by transdifferentiation of
mature hepatocytes and BECs rather than bipotent liver progenitor cells in mouse models [93, 101, 130, 134-137].

The first scRNA-seq studies to analyze purified BECs were conducted in mice and focused on BEC heterogeneity in homeostasis and following injury [45, 116]. In one study, 1,468 EPCAM+ BECs from two C57Bl/6 mice subjected to 16 days of 3,5-diethoxycarbonyl-1,4-dihydrocololidine (DDC) feeding were analyzed and characterized into three major categories: proliferating BECs, inflammatory BECs, and BECs expressing hepatocyte-related genes [116]. Classical BEC markers, including Epcam, Spp1, and Krt19 were noted to be expressed in all cells. High levels of YAP target genes were also broadly expressed, but WNT pathway components, including Lgr5, were not [116]. Dimensionality reduction, a computational method that attempts to group cells based on their transcriptomes, failed to clearly resolve the well-defined “clusters” generally associated with distinct cell types in scRNA-seq analyses. While heterogenous distribution of genes associated with the three identified BEC subpopulations was observed, the lack of distinct clustering may suggest that transcriptomic differences between these populations are subtle. Notably, the only BECs analyzed in this study were isolated from damaged IHBDs undergoing ductular reaction, and matched control samples were not assessed. An independent study, published concurrently, examined homeostatic IHBDs by scRNA-seq on 2,344 EPCAM+ BECs isolated from three C57Bl/6 mice. In these analyses the authors described two major BEC subpopulations, based on low or high expression of YAP target genes, including Cyr61, Klf6, and Hes1 (also a Notch target gene) [45]. A subsequent experiment in the same study examined 1,268 BECs from a single mouse subjected to DDC for one week. While control BECs failed to form well-defined clusters on dimensionality reduction, control and DDC-treated BECs demonstrated strong differential clustering, explained in part by broad upregulation of inflammatory genes [45]. Increased YAP activation was also observed in DDC samples, potentially explaining why YAP heterogeneity was not observed in DDC-treated BECs from Planas-Paz, et al [45, 116]. The lack of distinct BECs “clusters” in both experiments is somewhat confounding in the pursuit of clear genetic biomarkers of BEC subpopulations. One possible explanation is that BEC transcriptomes are largely homogenous. However, technical nuances in scRNA-seq cell isolation, library prep, and sequencing depth can lead to unintended exclusion of difficult to isolate cell types or an inability to detect low-expressed genes, impacting the ability to detect or resolve biologically relevant cell subpopulations [138].

In a recent study building on their prior liver scRNA-seq atlas, Andrews, et al. applied single nucleus RNA-seq (snRNA-seq) in addition to scRNA-seq, to increase detection of rare cell types in human liver [133, 139]. snRNA-seq, which involves directly releasing nuclei from intact tissues, was compared to scRNA-seq and shown to improve representation of BECs as a percent of total nuclei/cells examined (3.4% BECs in snRNA-seq vs. 2.4% BECs in scRNA-seq, whole liver) [139]. With snRNA-seq, 6 clusters were identified among 448 BECs by SNN-Louvain clustering, a standard method applied to identify distinct subpopulations in single cell transcriptomics [139, 140]. BEC subpopulations were named “Chol-1” through “Chol-6” and classified as progenitor-like, cholangiocyte-like, and hepatocyte-like, based on relative expression levels of genes associated with each classification group [139]. Chol-4 BECs were classified as the most “mature” BEC subpopulation, based on the highest expression of canonical BEC markers like AQP1, KRT7, and KRT8. Similar to findings by Aizarni, et al, Chol-2, Chol-3, and Chol-6 were described as hepatocyte-biased, based on expression of hepatocyte-associated genes, HNF4A and ALB [130, 139]. Two populations, Chol-1 and Chol-5 were only well-represented in snRNA-seq datasets and, interestingly, both were characterized as a less mature/progenitor phenotype. Chol-1 was classified as the least differentiated progenitor, based on elevated expression of transcription factors associated with
BEC development and pseudotime computational algorithms, which attempt to predict cellular lineage relationships based on gene expression patterns between cell clusters [139]. Chol-5 was found to have the highest level of BCL2 expression and was therefore proposed to contain small BECs [139, 141]. Finally, single cell transcriptomic assays in human BECs contradicted mouse studies in that WNT signaling was among enriched pathways terms in GO analysis of Chol-4 and Chol-5 BECs, though WNT enrichment was observed only in scRNA-seq and not in snRNA-seq data [45, 116, 139]. Collectively, emerging single cell assays demonstrate that there is significant transcriptomic heterogeneity in BECs of the IHBDs, though the biological significance of interpopulation gene expression differences and similarity between mouse and human BEC heterogeneity remains unclear.

**Limitations of single cell transcriptomics in BEC biology**

While single cell transcriptomics represents a powerful approach to dissecting cellular heterogeneity, inherent limitations in both wet lab and computational aspects of scRNA-seq present unique challenges to examining cell types with poor a priori definitions of heterogeneity. As shown in Andrews, et al. and demonstrated independently in other cell/tissue systems, isolation and enrichment protocols can bias which cell types are observed in scRNA-seq experiments and even alter gene expression [139, 142, 143]. Differences between the number of cells analyzed, sequencing depth, and number of unique genes/transcripts detected per cell can all vary between labs as well as experimental “batches” in the same lab, and lead to difficulty in resolving conflicting observations (current best practices for single cell transcriptomic experiments are reviewed comprehensively in [138]). Depth of read and transcriptsgenes detected per cell, which are closely linked, can have particularly significant impacts on cell subpopulation resolution in tissues like IHBDs, where biologically significant gene expression differences are unlikely to arise from highly expressed (and easily detected) genes [144]. While computational resolution of distinctly different cell types (like mesenchymal vs. epithelial cells) is facile using standard scRNA-seq analysis pipelines, identifying unique subpopulations in a relatively homogenous tissue remains challenging. Dimensionality reduction and clustering algorithms are also highly sensitive to user input, and decisions about clustering “resolution” (i.e.: how many clusters are identified by computational pipelines) often require prior knowledge of subpopulation biomarkers or are somewhat arbitrary [145, 146]. Even the task of comparing previously defined cell subpopulations between different scRNA-seq datasets remains challenging, with a growing number of computational tools being produced to “match” clusters generated in different experiments, often through machine-learning based approaches [147, 148].

Another limitation of scRNA-seq studies is the loss of spatial information, making it difficult to frame novel BEC subpopulations in context of classical small vs. large BEC definitions. While emerging spatial transcriptomic assays have been applied to liver sections in several studies, currently accessible commercial solutions for spatial transcriptomics fail to reach single cell resolution [139, 149]. Since BEC nuclei tend to be tightly packed within sections of IHBDs, these assays are likely insufficient to capture BEC-BEC heterogeneity. Further, the complex, three-dimensional structure of IHBDs presents an additional hurdle to spatial transcriptomic analyses, as many BEC-BEC spatial relationships may be hard to capture on tissue sections. Advances in spatial transcriptomic assays that facilitate single cell/single molecule resolution, as well as three-dimensional imaging, will be of great interest in further dissecting BEC heterogeneity [150, 151].

**Lineage relationships and cellular identities of transcriptomically-distinct BEC subpopulations**
Pseudotime analysis of human BEC scRNA-seq in Andrews, et al. was used to identify a putative progenitor subpopulation of BECs, and a similar approach was employed to argue that EPCAM+/TROP2\textsuperscript{int} cells represent bipotent progenitors in human liver [130, 139]. While pseudotime analyses are a commonly used computational tool for predicting lineage relationships between subpopulations identified by scRNA-seq, their accuracy remains controversial [152, 153]. The determination of a BEC progenitor cell type in adult IHBDs is also somewhat difficult to interpret in light of known BEC biology. In mouse and rat models, BEC proliferation at homeostasis has been reported to be <5% by immunohistochemical detection of proliferative markers KI67 or PCNA, with some studies reporting <1% proliferating BECs [92, 96, 154, 155]. Studies of proliferative heterogeneity during TAA induced ductular reaction identified highly proliferative BEC clones in small ductules, while BDL is associated with proliferation in large ducts [96, 156]. While this could indicate the existence of a singular progenitor-like BEC population interspersed throughout the IHBDs, it could also suggest that multiple populations of functionally distinct BECs (i.e. small and large BECs) share similar proliferative capacity after injury and respond in context-dependent manners.

An alternative possibility for relationships between BEC subpopulations observed in scRNA-seq data is that individual BECs cycle through diverse transcriptomically-defined “states”, rather than existing as fixed cell “types” defined by stable gene expression. It would then be possible for BECs to adopt new transcriptomic states without proliferating or giving rise to new cells. This model is supported in part by studies from Pepe-Mooney, et al. demonstrating that BECs cycle through transcriptional states defined by YAP levels [45]. An elegant lineage tracing approach demonstrated that BECs could cycle from YAP\textsuperscript{hi} to a YAP\textsuperscript{lo} state, by using Cre-\textit{lox} and smFISH to genetically label Yap\textsuperscript{hi} cells and examine transcription of Yap target genes over time. These experiments showed that BECs labeled while they were expressing high levels of YAP\textsuperscript{hi} marker \textit{Hes1}, could later downregulate YAP targets, including \textit{Cyr61}, in the absence of mitosis. YAP activity was further shown to be upregulated in response to bile acids through ASBT, suggesting that dynamic BEC transcriptional programs are responsive to environmental changes [45]. In further support of Yap-mediated BEC heterogeneity, Sox9\textit{EGFP} levels in BECs were shown to correlate with expression of Yap target genes, with EGFP\textit{lo} cells being consistent with Yap\textit{lo} BECs and EGFP\textit{hi} cells demonstrating increased expression of Yap target genes [102]. While these data present a compelling argument for dynamic BEC “states”, it remains unknown if observed changes in YAP target genes reflect similar dynamics in global, transcriptome-level expression changes.

With the emergence and progressively lower cost barrier of scRNA-seq, one immediate solution to circumvent the lack of consensus BEC subpopulation biomarkers is to apply unbiased single cell transcriptomics to studies of IHBD biology. A recent study on the role of primary cilia in hepatic cyst formation leveraged this approach to demonstrate that loss of \textit{Wdr35} in adult BECs upregulates TGF-\beta and extracellular matrix genes to drive cystogenesis in IHBDs [157]. In this case, single cell transcriptomics allowed for consideration of BEC heterogeneity while facilitating mechanistic insight into a disease phenotype.

**Conclusions and outlook**

Despite known BEC morphological and functional heterogeneity and several candidate biomarkers (ST14, Yap activity, Sox9\textit{EGFP} levels), a widely accepted consensus set of genetic biomarkers that can be used to identify discrete BEC subpopulations remains elusive. The ability of the field to establish such
a gene set from existing knowledge may be hindered in part by species-specific differences between models used to study BECs and the ability of such models to recapitulate human liver biology. Emerging evidence also suggests that the intrinsic biology of BECs themselves may defy simple classification into fixed subpopulations. In all currently published scRNA-seq datasets, expression of genes identified as cluster-specific markers does not appear to be unique to any one subpopulation of BECs. Rather, genetic heterogeneity of BECs appears to emerge largely from differences in expression levels of gene sets, rather than an “on”/”off” state for individual biomarkers [45, 116, 130, 139]. Reaching a “ground truth” definition of different BEC subpopulations might therefore require novel conceptual framework that moves beyond the search for single, cell type defining biomarkers often applied in other epithelial tissues. Additionally, if individual BECs are capable of actively cycling between distinct transcriptomic states, the task of assigning BECs to subpopulations representing stable cell identities might be conceptually impossible. Further studies are needed to generate a deeper understanding of BEC heterogeneity, dynamic transcriptional changes in IHBDs, and how BEC states or types may change in the setting of disease or injury.
Figure Legends

Figure 1. Heterogeneity within small and large intrahepatic biliary epithelial cells. The apical membrane of each BEC lines the ductal lumen within which melatonin and bile acids, large BEC-specific signaling molecules, are found. Large ducts have a larger diameter in the ductal lumen than small ducts. When compared to small BECs, large BECs are more columnar in morphology and have a reduced nucleus to cytoplasm ratio. In small BECs, NFAT2, NFAT4, and SP1 are transcription factors that localize to the nucleus. Also, in small BECs, HRH1 localizes to the basolateral membrane. Large BECs express MTR1/2, ASBT, AE2, and CFTR on the apical membrane and SCTR, HRH2, and SSTR on the basolateral membrane.

Table 1. Comparative liver anatomy across commonly used models of intrahepatic biliary biology.

Table 2. Defining characteristics between rat small and large intrahepatic bile ducts. Measurements provided for duct diameter, cell size, and cilia length are from rat IHBDs.
Acknowledgments

The authors would like to thank Dr. Fransky Hantelys for critical reading of the manuscript and helpful comments.
References


Xu WP, Cui YL, Chen LL, Ding K, Ding CH, Chen F, Zhang X, Xie WF. Deletion of Sox 9 in the liver leads to hepatic cystogenesis in mice by transcriptionally downregulating Sec 63. The Journal of Pathology 2021.


**Zebrafish**
- Lobation: Bilobated; Dorsal lobe (DL) and ventral lobe (VL)
- Gallbladder (G): Present
- Lobule organization: Absent
- Portal triad: Intermittently dispersed
- Hepatic artery: Poorly identified/unidentified
- Hepatocyte zonation: Absent
- Intrahepatic bile duct heterogeneity studies: Small and large ducts, scRNA-seq

**Rat**
- Lobation: Lobated; Left lobe (LL), median lobe (ML), right lobe (RL), caudate lobe (CL)
- Gallbladder (G): Absent
- Lobule organization: Present
- Portal triad: Present
- Hepatic artery: Present
- Hepatocyte zonation: Present
- Intrahepatic bile duct heterogeneity studies: Small and large ducts, scRNA-seq

**Mouse**
- Lobation: Lobated; Left lobe (LL), median lobe (ML), right lobe (RL), caudate lobe (CL)
- Gallbladder (G): Present
- Lobule organization: Present
- Portal triad: Present
- Hepatic artery: Present
- Hepatocyte zonation: Present
- Intrahepatic bile duct heterogeneity studies: Small and large ducts, scRNA-seq

**Human**
- Lobation: Not lobated; However, the human liver is described in segments of left, right, and caudate (not shown) lobes
- Gallbladder (G): Present
- Lobule organization: Lobated; Left lobe (LL), median lobe (ML), right lobe (RL), caudate lobe (CL)
- Portal triad: Present (Same as above)
- Hepatic artery: Present
- Hepatocyte zonation: Present
- Intrahepatic bile duct heterogeneity studies: Small and large ducts, scRNA-seq, snRNA-seq

**Citations**
- [37, 48]
- [21, 23, 45, 52, 96, 102, 111, 116, 130, 131, 133, 139, 140]
<table>
<thead>
<tr>
<th>Duct diameter</th>
<th>Small IHBDs</th>
<th>Large IHBDs</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;15 μm</td>
<td>8.49 ± 0.05 μm</td>
<td>14.53 ± 0.05 μm</td>
<td>[23]</td>
</tr>
<tr>
<td>&gt;15 μm</td>
<td>7.28 ± 1.32 μm</td>
<td>7.28 ± 1.32 μm</td>
<td>[66, 67]</td>
</tr>
<tr>
<td>Cell size</td>
<td>HRH1, NFAT2, and NFAT4</td>
<td>SCTR(SR), Cl-HCO3-</td>
<td>[21, 76-91]</td>
</tr>
<tr>
<td>Cilia length</td>
<td>HRH1, NFAT2, and NFAT4</td>
<td>AE2, CFTR, ABAT, NFAT3, SSTR, MT1, and MT2</td>
<td>[76-80, 79, 86, 94]</td>
</tr>
<tr>
<td>Markers</td>
<td>cAMP dependence; HRH2 regulation; proliferate in response to BDL</td>
<td>[35, 90, 144]</td>
<td></td>
</tr>
</tbody>
</table>