HUMAN STEM CELL MODELS IDENTIFY TARGETS OF HEALTHY AND MALIGNANT HEMATOPOIETIC REGULATION

BY

JENNIFER C REID, BSc

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

McMaster University

Descriptive Note

McMaster University BACHELOR OF SCIENCE (2015) Hamilton, Ontario (Biomedical Research Specialization, Co-op)

Title: Use of human stem cell models to identify targets of healthy and malignant hematopoietic regulation

Author: Jennifer C. Reid

Supervisor: Dr. Mickie Bhatia

Number of pages: xii, 163

Abstract

Hematopoiesis is the highly regenerative process of producing billions of blood cells each day, including white blood cells, red blood cells, and platelets. Given the relatively short life span of these mature cells, hematopoiesis is dependent on stem and progenitor cells to generate renewed progeny, which represents a tightly regulated process. This includes cell intrinsic and external factors, and where dysregulation can lead to anemia and cancer. As such, the hematopoietic hierarchy has been intensely studied for nearly a century and represents a gold standard model of cell fate and developmental biology, in research and clinical applications. Cellular models, such as *in vitro* culture and human-mouse xenografts *in vivo*, have been developed to explain complex phenomena pertaining to hematopoiesis and also interrogate processes which are too invasive to study in humans. Hematopoietic generation is required beyond sustaining homeostasis, and progenitors can be damaged through cytotoxic injuries such as radiation and standard chemotherapy, and also undergo leukemic transformation. There are two main treatment modalities for leukemia patients (a) receiving a stem cell transplant, and (b) drug or radiation-based therapy. In the former, shortages of donors and stem cells has remained an unmet clinical need for decades. In the latter, selective targeting of genetic mutations has become a successful standard-of-care in leukemias such as chronic myelogenous leukemia and acute promyelocytic leukemia. However, in the most common adult hematologic malignancy, chronic lymphocytic leukemia (CLL), similar targeting therapies have not been developed. Altogether, shortages of stem cells from healthy donors, chemotherapy-induced immune dysfunction, and a lack of targeted therapies, all reinforce the immediate need for innovative cellular models to address these clinical problems.

To generate additional sources of human hematopoietic progenitors for laboratory study, human PSCs have been used. Unlike hematopoietic progenitor cells collected from healthy and leukemic donors, human pluripotent stem cells (PSC) can be easily propagated and expanded *in vitro*. PSCs can generate hematopoietic progenitor cells, but they remain poorly understood and have not been robustly applied to solve the aforementioned deficiencies related to patient treatment. Importantly, the biological regulation of both hematopoiesis and PSCs has been experimentally confirmed to significantly deviate between humans and other animals, such as mice, further reinforcing the importance of human-specific cell models of hematopoiesis. *Therefore, I hypothesized that human stem cell models provide a focused approach to interrogate the regulation of hematopoiesis from the apex of the hierarchy, which can be used to understand the promotion of healthy hematopoiesis and understand malignant transformation.*

We began by providing evidence that a discovery platform based on human PSCs with a specific leukemic driver mutation, trisomy 12 (tri12), enabled unprecedented elucidation of cell autonomous pathways for precision oncology, specifically in the context of tri12 CLL. Together with traditional gene expression analysis, we also developed machine learning algorithms to molecularly dissect pathways specific to tri12 in PSC and CLL, that could only be found using human isogenically-controlled PSC with a genetic leukemic driver *in vitro*. Two actionable targets were therapeutically applied to target tri12 CLL patient samples and tri12 PSCs both *in vitro* and upon xenotransplantation as tumors into immunocompetent mice.

We then investigated what is necessary to improve PSC-derived hematopoietic progenitors to extend their application beyond in vitro models, and also to address the clinical shortage of transplantable stem cells. We discovered that PSC-derived progenitors fail to survive even 24 hours following bone marrow transplantation, while they can survive and proliferate for weeks *in vitro*. We then determined PSC-derived progenitors do not have a key receptor that healthy progenitors have, called CXCR4. By correcting CXCR4 signaling, we observed enhanced PSC-derived progenitor function and survival in the bone marrow of human-mouse xenografts and a transcriptional shift toward human progenitor profiles. Our results reveal CXCR4 is an actionable target for transplantable progenitors with *in vivo* function.

To investigate hematopoietic dysfunction, which is a pervasive side-effect of radiation and chemotherapy *in vivo*, we developed a third human cellular model using long-term human-mouse xenografts given these broad effects cannot be recapitulated *in vitro*. We sampled xenografts at multiple time points before and during the recovery process after treatments and observed drastic yet reproducible changes in the composition of human progenitors and also mature human hematopoietic cells, which reflected clinical leukemia patient experiences. Furthermore, using single cell transcriptional and protein analysis, we were able to identify unique subsets of human progenitors which may prove to be the key hematopoietic regenerating cells upon further biological investigation of these actionable targets. Validation of clinically relevant interventions which aid in the process of healthy hematopoietic regeneration *in vivo* are the next step in this work. Collectively, the data presented within this thesis offer a deeper conceptualization of human stem cell models and the deconvolution of several complex components of hematopoietic regulation. This work has revealed novel, clinically relevant, and actionable targets to ultimately enable the promotion of healthy hematopoiesis on multiple fronts.

Acknowledgements

Mick, I want to thank you for all the once-in-a-lifetime memories; from litterbox and oompa loompas to creative brain storming and motivational speeches. I have learned so much from you, and I am grateful for that. You have taught me so much, and also let me get messy and learn from my mistakes. I appreciate that a thesis should capture the 'hard skills' and data generated, but I would be amiss not to expressly thank-you for the even longer list of 'soft skills' and personal development that are not directly mentioned in this thesis.

I'd also like to thank my committee members, Drs. Thomas Hawke and Greg Steinberg, who have asked out-of-the-box questions and advised me during my graduate studies. I have admired your character since my undergrad days, and I am so grateful to have such positive roles models.

My graduate success would not have been possible without my lab family. Diana and Allison, you have become true friends and I am so grateful for everything we've shared. I want to thank Jong-Hee and Kyle for taking a chance on me and trusting me to articulate and publish their work. I want to thank everyone whom I have had the pleasure to work with, including current members and alumni of the Bhatia lab; JB, Borko, Macky, Luca, Deanna, Gena, Yeonjoon, Juan, Andrew, Lili, Mio, and the bhatiabhaddies; you make my day.

To all the brave patients, and their families, who participated in clinical trials and donated valuable tissue samples used in this and other studies, I am truly grateful.

Justin, whether you're cooking me dinner or at the gym while I'm at the lab, I wouldn't have been able to finish this degree without your love and support. Here's my chance to publicly apologize for every time I was late leaving the lab.

Last, but not least, I want to thank my family, who continuously promote me to become a better version of myself.

❖

"There is no point in starting something without ambition. ... Ambition never goes away. It may shuffle off, grumbling, feet dragging, only to slide across into something else – usually the next project. It doesn't take 'no' for an answer. ... One last word to all you nascent writers out there. Ambition is not a dirty word. Piss on compromise. Go for the throat. Write with balls, write with eggs. Sure, it's a harder journey but take it from me, it's well worth it. – Steven Erickson [Steve Rune Lundin], 2007

❖

Table of Contents

List of Figures and Tables

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

List of Abbreviations

3TF	3 transcription factors (*HOXA9*, *ERG*, and *RORA*)
5TF	5 transcription factors (*HOXA9*, *ERG*, *RORA*, *SOX4*, and *MYB*)
ANOVA	Analysis of variance
AraC	Cytarabine
bFGF	Basis fibroblast growth factor
BCL2	B cell leukemia 2
BM	Bone marrow

BPM4	Bone morphogenetic protein 4
BTK	Bruton's tyrosine kinase
CB	Cord blood
CFU	Colony forming unit
CFU-S	Colony forming unit - spleen
CIM	Chemotherapy-induced myelosuppression
CLL	Chronic lymphocytic leukemia
CX3CL1	C-X-3-C ligand 1
CX3CR1	C-X-3-C receptor 1
CXCL12	C-X-C ligand 12
CXCR4	C-X-C receptor 4
ECC	Embryonal carcinoma cells
EDNRB	Endothelin receptor B
ESC	Embryonic stem cell
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FDR	False detection rate
FISH	Fluorescence *in situ* hybridization
FLT3L	Fms-related tyrosine kinase 3 ligand
G-CSF	Granulocyte-colony stimulating factor
GEO	Gene expression omnibus
GFP	Green fluorescence protein
GM-CSF	Granulocyte-macrophage colony stimulating factor
GSEA	Gene set enrichment analysis
HBSS	Hank's buffered salt solution
HEK	Human embryonic kidney
HPC	Hematopoietic progenitor cell
HRC	Hematopoietic regenerating cell
HSC	Hematopoietic stem cell
HSCT	Hematopoietic stem cell transplantation
HSPC	Hematopoietic stem and progenitor cells
IF	Intrafemoral
IGHV	Immunoglobulin heavy-chain variable region
IL2Rγ	Interleukin 2 receptor, gamma subunit
IL3	Interleukin 3
IL6	Interleukin 6
IL7	Interleukin 7
IMDM	Iscove's modified Dulbecco's medium
IP	Intraperitoneal
iPSC	Induced pluripotent stem cells
IRES	Internal ribosome entry site
IV	Intravenous
KO-DMEM	KnockOut-Dulbecco's Modified Eagle Media
Lin^-	lineage-depleted
MEFCM	Mouse embryonic feeder-conditioned media

MOI	Multiplicity of infection
MPB	Mobilized peripheral blood
MNC	Mononuclear cell
MPB	Mobilized peripheral blood
NES	Normalized enrichment score
NGS	Next generation sequencing
NOD/SCID	Non-obese diabetic/Severe combined immunodeficiency
NOM	Nominal
NSG	NOD/SCID/IL2Rγ^{null}
PB	Peripheral blood
PCA	Principal components analysis
PCR	Polymerase chain reaction
PMID	PubMed identifier
PSC	Pluripotent stem cell
RNAseq	Ribonucleic acid sequencing
SCF	Stem cell factor
SCID	Severe combined immunodeficiency
scRNAseq	Single cell ribonucleic acid sequencing
SE	Standard error
SEM	Standard error of the mean
SNP	Single nucleotide polymorphisms
SQ	Subcutaneous
SRC	SCID repopulating cell
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
tri12	Trisomy 12
t-SNE	t-Distributed stochastic neighbor embedding
Tx	Treatment
UCB	Umbilical cord blood
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier

Declaration of Academic Achievement

I contributed to the design and execution of the research presented, as well as performed data analysis and writing of all sections of this thesis. Dr. Mick Bhatia directed the research projects, aided in the analysis of data, and editing of this thesis. Additional specific contributions by other lab members and external collaborators are noted in Chapters 2 through 4.

Chapter 1: Introduction

1.1 Overview

This thesis presents research on novel molecular and genetic regulatory pathways of self-renewal and differentiation in models of healthy and malignant human hematopoiesis. The origin of healthy hematopoietic regulation stems from a large body of work spanning decades and encompasses many efforts by others to derive HSCs from human pluripotent cells. The development of a genetic model for the malignant regulation of CLL was truly serendipitous, was propelled through robust and intriguing results that begged for further exploration, and filled a clinical gap in identifying actionable targets in CLL. Lastly, these two projects, along with my supportive roles in other published works throughout my graduate studies, instructed me to develop an *in vivo* model to uncover the biology of regenerating healthy hematopoiesis during injury. The goal of this introductory chapter is to briefly introduce the history of HSCs and PSCs, using healthy and malignant scenarios, and summarize the current state of these research fields. This thesis was completed exclusively using human HSCs and PSCs, as such, ***Chapter 1*** will heavily focus on basic human biology. However, in most cases, knowledge of the human system was preceded and instructed by the murine system. This will be dully acknowledged throughout but not thoroughly detailed. This chapter closes with the description of multiple bioinformatic tools and descriptions necessary to appreciate the complexity and opportunity available in the 'Information Age' of big data, as it applies to HSCs and PSCs.

1.2 Healthy hematopoietic stem cells

The self-renewal and differentiation of multipotent hematopoietic stem cells (HSCs) is a highly controlled process that occurs over a lifetime, where HSCs are the sole source of the continuous supply of billions of mature blood cells produced every day (Doulatov, 2012). Exhaustion or damage of HSCs can lead to anemia, cytopenia, and malignant transformation. Hematopoietic progenitor cells (HPCs) are progeny of HSCs with reduced self-renewal and increased proliferation. Hematopoietic stem and progenitor cells (HSPC) are sometimes referred to collectively when indicating primitive hematopoietic cells without making claims of their specific functional capacity.

1.2.1 Obtaining hematopoietic stem and progenitor cells

The hematopoietic system of human newborns transitions from the liver to the highly vascularized and innervated tissues encased in solid bone, which is called bone marrow (BM) (Mikkola and Orkin, 2006). Studying liquid tissue encased in solid bone presents an inherent challenge. Collecting donor HSCs therefore requires either extracting them directly from within bones or releasing them from the BM into circulation for venous peripheral blood collection. The former requires trephine biopsy needles and is usually taken from the pelvic bone by a hematologist. The latter is typically achieved with mobilizing agents that disrupt the cellular adhesion of HSPCs and permit their egress into circulation; mobilized peripheral blood (MPB) (Bain, 2001; Petit et al., 2002). Questionably, early researchers suggested chemotherapy as a treatment to mobilize HSCs into circulation to improve collection yields

(Richman et al., 1976). However, less toxic procedures are now the standard of care. As such, mobilization via granulocyte-colony stimulatory factor (G-CSF) leads to the degradation of cellular adhesion molecules tethering HSPCs to the BM increasing their presence in peripheral circulation (Dührsen et al., 1988). A third common source of HSPCs is from umbilical cord blood (CB) (Broxmeyer et al., 1989), which is more highly enriched in functional HSCs than the former two sources (Wang et al., 1997). These three sources represent the only access to clinical-grade HSCs.

1.2.2 Homeostasis and injury

Homeostatic and injury states of hematopoiesis are both regenerative in nature. Homeostatic hematopoiesis requires the balanced generation of a trillion blood cells every day (Doulatov, 2012). However, diverse insults can change and dysregulate this balance in cases of infection, inflammation, hematological cancers, medically necessary chemotherapy, and radiation (Batsivari et al., 2020). Injury also changes the composition of non-hematopoietic BM cells and the chemical signals they express to support hematopoietic recovery and regulate HSCs (Silberstein et al., 2016; Tikhonova et al., 2019). Early responses following acute injury recruit the proliferation and differentiation of HSCs (Mauch et al., 1995), as well as non-hematopoietic BM cells (Tikhonova et al., 2019). Transient expansion of the progenitor compartment also supports emergency production of BM and blood cells, known as *myelopoiesis* (Herault et al., 2017). However, incomplete resolution of injuries persist (Milyavsky et al., 2010) and may lead to secondary malignancies even when the original therapy appeared curative (Mauch, 1995). More recent studies have found that stressed or damaged HSCs repair more slowly than progenitor cells (Milyavsky et al., 2010), will activate programmed cell death and have impaired repair mechanisms, which can lead to malignant transformation and long term accumulation of mutations (Mohrin et al., 2010). This contrasts with progenitors and most mammalian cell types which will seek to repair themselves efficiently. Repeated exposure to injury induces *HSC exhaustion*, leading to long term deficits in balanced hematopoiesis (Zhang et al., 2016), which is experimentally determined in mice (more in *Section 1.2.4*).

1.2.3 In vitro *assays of self-renewal*

In order to follow the differentiation of individual clones of hematopoietic cells, semi-solid media substrates were developed *in vitro* which could hold progenitors "*in situ*" while they proliferated (Ichikawa et al., 1966). Mouse BM cells would be plated in semi-solid media with growth factors, and the ensuing frequency of colony formation quantitatively identified the number of progenitor cells that were plated. This *in vitro* format of progenitor quantitation was readily adapted to human hematopoietic cells (Pike and Robinson, 1970). The lineage contribution and morphology of colonies from these early studies demonstrated that myelo-erythroid lineage cells were more readily generated, such as monocytes and macrophages, granulocytes, and precursors to red blood cells (Ichikawa et al., 1966), and further refinement permitted the study of megakaryocytic progenitors (Kimura et al., 1984). However, there is a paucity of *in vitro* assays testing T and B cell clonogenic potential (Punzel et al., 1999), making the hierarchy of all progenitors, and testing for potential lineage bias, difficult to demonstrate *in vitro*.

1.2.4 In vivo *assays of self-renewal*

Evidence for the existence of hematopoietic stem cells (HSC) occurred during the atomic age, where the cause of radiation-induced death was found to be BM failure (Jacobson et al., 1951). Strikingly, lethally irradiated animals could be rescued by providing allogeneic BM transplantation after irradiation (Lorenz et al., 1951). Early studies also demonstrated, using distinct chromosomal markers, that it was the donor cell that reconstituted the host's hematopoietic system (Ford et al., 1956), rather than repairing the damaged cells via circulating factors (Jacobson et al., 1951). None of these studies demonstrated that a single multipotent HSC existed. However, this changed when Canadians, Till and McCulloch, demonstrated that single clonogenic BM cells were able to self-renew and restore hematopoiesis when transplanted into lethally irradiated recipient mice (Becker et al., 1963). They identified nodules of cells forming on the recipients' spleens, which were proportional in number to the original number of BM cells injected, and termed colony forming unit-spleen (CFU-S). Further work demonstrated multilineage hematopoiesis within individual CFU-S clonal nodules (Wu et al., 1968), and in doing so, mouse HSCs were functionally demonstrated to exist.

Around this time, human hematopoietic stem cell transplantation (HSCT) was successfully applied to replenish the hematopoietic system of cancer patients undergoing radiation and chemotherapy who also had an identical twin (Thomas et al., 1959). Accidental exposure of humans to lethal radiation was quite rare and patients did not survive, despite HSC transplants, due to graft vs. host disease (Mathé et al., 1959a; Mathé et al., 1959b). This precluded the experimental study of human HSCs until a suitable surrogate transplant recipient could be developed. Therefore, the discovery of the severe combined immunodeficiency (SCID) spontaneous mutation was an opportune windfall (Bosma et al., 1983). Although the original authors modestly recommended their SCID mouse as a model of impaired T and B cell differentiation, its greatest contribution has been in human-mouse xenotransplantation studies, which lead to further immune compromised mouse strains (Ito et al., 2012). Indeed, primitive hematopoietic cells from MPB, BM, and CB could lead to multilineage engraftment of preconditioned immunocompromised mice, and were termed SCID repopulating cells (SRC) (Bhatia et al., 1997; Wang et al., 1997). Today, the NSG mouse with a **N**on-obese diabetic (NOD) background, the **S**CID mutation, and a IL2 receptor **g**amma deletion represents a better recipient of human tissue for engraftment to date, lacking innate and adaptive immunity components (Ito et al., 2012). For uniformity with the current terminology of the field, this thesis refers to SRC as HSC, despite the functional definition of human HSCs as being able to repopulate humans.

1.2.5 Purification of hematopoietic stem and progenitor cells

Approximately 1 in a million human bone marrow cells is a functional HSC (Wang et al., 1997), requiring purification from the rest of the more differentiated cells in order to study their functional and molecular properties. Human HSCs are enriched by positively selecting for CD34 and CD90 cells and by negatively selecting for mature markers, such as CD38. However, the stage of ontogeny also appears to influence HSC phenotype as fetal HSCs preferentially express the endothelial protein C receptor while umbilical cord blood (UCB) HSCs preferentially express integrin α6, also known as CD49f (Subramaniam et al., 2018).

Some markers have functional significance, such as mediating cellular adhesion in a microenvironment (Healy et al., 1995; Subramaniam et al., 2018). To date, the complete phenotype of human HSCs has not been described, thereby hindering the absolute purification of human HSCs (Baron et al., 2019). In contrast, murine HSCs have been isolated with a cocktail of six cell surface markers and the absence of mature lineage markers (Lin-) (Oguro et al., 2013). Identification of novel markers to purify single HSCs using cell surface markers would provide a technological and biological advance, for example, in studying the cell fate decisions of aging HSCs, or studying leukemic alterations and injury responses. Identification and biological validation of markers from single cell RNA sequencing data sets (*more on this subject in Section 1.6*) enriched for hematopoietic progenitors (Pellin et al., 2019), and others, are likely to provide actionable targets and markers for further characterization.

1.3 Pluripotent stem cells

Renowned for their ability to give rise to all tissue types of the body, as well as unlimited self-renewal capacity, human pluripotent stem cells (PSC) are an exciting tool with wide-ranging applications in understanding early human development, regenerative medicine, and disease modelling. In this section, I will review the road which lead us to human PSCs, gold standard measurements of pluripotent potential, as well as highlighting some key applications of human PSCs to date.

1.3.1 Sources of pluripotent stem cells

Foundational observations which fueled the field of PSC research occurred through serendipity. An inbred strain of mice known as "129" frequently presented with spontaneous testicular tumours, which permitted their study (Stevens and Little, 1954). Upon histological examination, it was discovered that these tumors were comprised of very many different tissue types, and that the arrangement of tissues appeared disorganized. These types of growths were termed *teratocarcinomas*; malignant tumours comprised of tissues derived from all three embryonic germ layers (ectoderm, endoderm, and mesoderm), which exist in a disorganized mass. As such, the name is rather fitting since "*teras*" is the Greek word for monster. The cells which give rise to teratocarcinomas originate from germ cells in sex organs (testis and ovaries), which are believed to have "gone rogue" and are technically referred to as pluripotent embryonal carcinoma cells (ECC) (Martin, 1981). The intentional generation of mouse embryonic stem cells (ESC) was achieved roughly two decades later (Evans and Kaufman, 1981; Martin, 1981), and the term "ESC" was introduced to distinguish these cells from ECCs (Martin, 1981). Mouse ESCs were derived through the destruction of mouse blastocysts and defining culture requirements which sustained an undifferentiated state *in vitro*.

The ability to derive human ESC took yet another decade of work to bear fruit and was also preceded by the derivation of non-human primate ESC by the same research group. Jamie Thomson, from the University of Wisconsin, selected non-human primate ESCs from the inner cell mass of blastocysts and plated them on supportive feeders cells which selected for ESC lines that could be propagated in an undifferentiated state for over a year of *in vitro* culture (Thomson et al., 1995). Three years later, Thomson "essentially" used the same

techniques to generate human ESC lines, which again could be serially propagated in an undifferentiated state while maintaining a normal karyotype (Thomson, 1998). Two of these human ESC lines in particular, H1 and H9 (also known as WA01 and WA09), are still in use today and have been adopted by our research group and by many groups worldwide (Nakanishi et al., 2019; The International Stem Cell Initiative et al., 2007; The International Stem Cell Initiative et al., 2011). While the original human ESC lines were derived from "poor" blastocysts and discarded from *in vitro* fertilization (IVF) clinics with parental consent, it ignited ethical debates and nationally-imposed limitations on human ESC research funding (Anonymous, 2010). However, less than a decade after the derivation of human ESCs, Japanese researchers demonstrated the ability to alter the cell fate of terminally differentiated somatic cells to become cells which closely resembled ESCs, which they termed *induced pluripotent stem cells* (iPSC) (Takahashi et al., 2007; Takahashi and Yamanaka, 2006). The use of iPSC completely circumvented the controversy regarding the destruction of human blastocysts, but it was the concepts and techniques developed through human ESC technology that enabled the creation of iPSC. Not only was iPSC technology rapidly adopted and robustly reproduced internationally, it achieved what was thought to be impossible; using only four transcription factors to change cell fate and "turn back the clock of development".

1.3.2 Measuring human pluripotent potential

What was striking about the spontaneous testicular tumours of strain 129 mice observed by Stevens and Little was that they exhibited "progressive growth", were serially transplantable, and were comprised of many different tissue types (Stevens and Little, 1954). The refinement of these features would become the gold standard functional assay for measuring pluripotent potential; *the teratoma assay* (Hentze et al., 2009). Teratoma formation is affected by multiple factors including the delivery route, type of PSC, and cell number. Currently, testing for pluripotent potential entails injecting the test cells into immunodeficient mice, with sites such as hind limb muscle, kidney capsule, testes, subcutaneous, and intraperitoneal spaces all permitting tumor formation (Hentze et al., 2009; Stevens and Little, 1954). Our group demonstrated the effects of different types of PSCs on the types of tissues which were observed in the resulting teratomas, as well as the dependence of tumour formation based on cell number (Werbowetski-Ogilvie et al., 2009). Upon resection of the tumor and histological preparation, tissues comprising all three germ layers must be observed (ectoderm, endoderm, and mesoderm) (Hentze et al., 2009). This suggests pluripotent precursor cells existed at the time of injection which were capable of giving rise to more differentiated tissues, which is known as *differentiation potential*; the first functionally defined key measure of pluripotent cells. The second measure of pluripotent cells is the capacity to produce progeny which maintain an equivalent differentiation potential, which is known as *self-renewal*. As key distinction is that while self-renewal requires cell division, cell division alone does not demonstrate self-renewal. To functionally demonstrate self-renewal, the PSC which gave rise to a teratoma must also be able to divide and maintain equivalent differentiation potential. This means that within the primary teratoma, at least one progeny cell exists which could give rise to another teratoma. This is experimentally demonstrated by excising the primary teratoma, mincing it into smaller pieces, then transplanting in into a second mouse to determine whether another teratoma forms (Stevens and Little, 1954).

While the primary teratoma has become the routine test to demonstrate pluripotent potential for ESCs (Evans and Kaufman, 1981; Martin, 1981; Nakanishi et al., 2019; Thomson et al., 1995) and iPSCs (Lee et al., 2017b; Takahashi et al., 2007; Takahashi and Yamanaka, 2006), routine testing for self-renewal by secondary transplant is uncommon.

1.3.3 Development of experimental models

Using mouse ESCs, contribution of pluripotent cells to the entire developing embryo can be achieved using tetraploid complementation. While human ESCs theoretically possess the potential to generate all tissue types of the human body, the international research community and ethical experts have agreed that contribution of human ESC to a developing human embryo is strictly prohibited. Indeed, human-mammal chimeras are typically prohibited in most jurisdictions. A notable – *ethically permitted* – exception to this general ban was conducted by researchers at the Salk Institute, who sought to improve the contribution of human cells into porcine embryos as a means to produce "humanized organs" for tissue replacement therapy (Wu et al., 2017). This was endorsed since organ xenotransplantation is of great interest to reduce organ donor shortages (Längin et al., 2018). However, even more closely related chimeric animals such as mouse-rat chimeras are naturally aborted. The human PSCs within human-pig chimeras were subject to random inclusion and experiments were terminated within 4 weeks (10% and 25% of human and porcine gestational length, respectively).

Arguably some of the best applications of human PSC experimental models have occurred *in vitro* when used for developmental biology and disease modeling. For example, it is not permissible nor feasible to study the origin of pluripotency in human sex tissues, and comparative biology approaches using mice have been inadequate. Therefore, our group studied the creation and maintenance of pluripotent material *in vitro* which infers upon developmental processes that remain unclear in human development *in situ* (Nakanishi et al., 2019). Human iPSC-derived cells also offer the ability to capture patient genetics, essential for disease modelling, which can then be differentiated towards tissues affected by the disease. Two successful examples of the power of human PSCs in disease modelling are demonstrating genetic causality in Hutchinson-Gilford progeria syndrome (Liu et al., 2011), and a study by our groups which associated a genetic aberration with a disease phenotype in leukemia (Lee et al., 2017b).

1.3.4 Hematopoietic differentiation of pluripotent cells

The ultimate goal of controlling differentiation of human PSCs is an intense field of research. Human PSCs spontaneously differentiate in culture in the absence of provided extrinsic growth factors, if they become too confluent or too sparse, or upon injection into immunodeficient mice (Thomson, 1998). While PSCs readily differentiate, the difficulty lies in coercing the differentiation towards the desired cell outcome and removing undesired cell types that may spontaneously form. This is known as *directed differentiation*. Many research groups have developed protocols which involve temporal administration of growth factors and small molecules, conditioned media, and/or feeder cells, which theoretically guide PSC along particular pathways of differentiation (Chadwick et al., 2003; Sturgeon et al., 2014; Wang et al., 2004). However, it remains unclear whether differentiation occurs

spontaneously, and desired cells are selected for survival by favourable media conditions. With the rapid expansion of labs using and differentiating human PSCs in the early 2000's, several groups, including ours, derived hematopoietic cells which were capable of CFU capacity and morphologically resembled somatic hematopoietic CFU (Chadwick et al., 2003; Vodyanik et al., 2005). This created demand for human PSC-derived HSCs which would theoretically solve shortages of donor material and the ability to create a "universal donor HSC" (Wang et al., 2005). Spontaneous hematopoietic development *in vivo* from human PSC-derived teratomas in mice was also observed (Amabile et al., 2013; Suzuki et al., 2013). However, this enthusiasm was tempered as SRC potential in mice was exceedingly rare or non-existent (Doulatov et al., 2013; Ran et al., 2013; Wang et al., 2005). Even with the most sophisticated protocol to date, human PSC-derived HSCs have not been created *in vitro* and appear to rely on as yet unspecified cues from murine transplant recipients (Sugimura et al., 2017). In this thesis, I will refer to human PSC-derived hematopoietic cells with progenitor capacity as such (HPC, hematopoietic progenitor cells), and not "HSPC", "HSC", or "HSC-like" as these are inaccurate; these cells lack SRC capacity.

What has become apparent across multiple reports, is the difficulty of B lymphoid cell development from human PSC sources (Kennedy et al., 2012; Vo et al., 2018; Vodyanik et al., 2005), where the highest efficiency reported to date for B lymphoid development is 2% or less from wildtype hPSCs (Carpenter et al., 2011; French et al., 2015; Galat et al., 2017). The issue of the paucity of B lymphoid development has not been addressed, which is striking considering the much more successful production of T lymphoid cells from hPSCs (Kennedy et al., 2012). Furthermore, while erythroid lineage cells can be produced, they default to express fetal hemoglobin instead of adult hemoglobin (McIntyre et al., 2013; Ran et al., 2013). These lines of evidence suggest that human PSC are producing hematopoietic cells that resemble those found in the human fetus, which is known as *primitive hematopoiesis* (Sturgeon et al., 2013). This distinction is important as HSCs do not sustain the hematopoietic system at this time. A switch occurs in neonates as HSCs drive *definitive hematopoiesis* from then until the end of life. In summary, researchers in the field have adapted differentiation programs to better understand the complexity of human hematopoietic development (Sturgeon et al., 2014), identifying deficiencies which must be overcome, step-by-step, to promote successful xenotransplantation (Lee et al., 2017a), as well as striving to produce SRC with somatic HSC efficiency. Despite these milestones, there is a paucity of evidence suggesting that human PSC-derived hematopoietic cells can interact at all with the microenvironment following xenotransplantation.

1.4 Chronic Lymphocytic Leukemia

1.4.1 Clinical characteristics, pathophysiology, and risk stratification

Chronic lymphocytic leukemia (CLL) is the most common adult hematologic malignancy. CLL is considered to be an indolent but incurable B-cell leukemia. Leukemic B cells undergo clonal expansion and share a well-defined immunophenotypic profile based upon CD5, CD19, CD20, and CD23 expression, as well as kappa or lambda (κ/λ) surface immunoglobulin light chain restriction (Parikh, 2018). CLL leads to acquired immune dysfunction and related complications in most patients (Parikh, 2018; The World Health Organization et al., 2014). Additionally, a subset of patients' disease will undergo transformation leading to Richter

Syndrome; an accelerated terminal phase of aggressive large cell lymphoma (Rossi et al., 2018). CLL is classically defined as monoclonal cellular expansion, which suggests genetic lesions could be driving disease manifestation. Genetic analysis indeed revealed that while CLL patients lack a single common molecular target in leukemic B cells, the disease is associated with a relatively small number of subgroups segregated by recurrent molecular lesions (Döhner et al., 2000). These recurrent abnormal chromosomal lesions became biomarkers, and have been incorporated into standard of care using fluorescence *in situ* hybridization (FISH) based detection (Parikh et al., 2016). FISH-based testing of chromosomal abnormalities strongly correlates to overall survival and can be segregated into five groups: deletion of chromosome 17 p-arm (del(17p)), del(11q), trisomy of chromosome 12 (tri12), normal karyotype, and del(13q) (in order from shortest to longest survival) (Döhner et al., 2000). Further characterization determined that leukemic B cells may arise from at least two different cellular origins. First, a mature B cell with re-arranged *IGHV* (immunoglobulin heavy-chain variable region) mutations, and secondly, from an earlier precursor, designated as *IGHV*-unmutated. All five FISH-based cytogenetic abnormalities have been observed in both *IGHV*-mutated and *IGHV*-unmutated cases, with the latter experiencing a worse prognosis upon diagnosis (Morabito et al., 2013).

Risk stratification for CLL patients began fifty years ago (Binet et al., 1981; Rai et al., 1975), which is largely based on the degree of tumor size and healthy hematopoietic suppression. Molecular classifications followed later (Döhner et al., 2000). One of the more recent and larger consortia on risk stratification in CLL published the "international prognostic index in chronic lymphocytic leukemia" (CLL-IPI). This stratified patients based on *TP53* and *IGHV* mutation status, serum β2-microglobulin concentration, disease state, and patient age at time of diagnosis (Bahlo et al., 2016). This risk stratification has been shown to be more predictive than other common prognostic indexes (Munoz-Novas et al., 2018). However, researchers have called for further stratification based on pathways and genetics, as an increasing proportion of patients are diagnosed with CLL-IPI low risk. CLL-IPI low risk does not provide clear direction and challenges prognostication, nor does it provide actionable targets for precision oncology (Brieghel et al., 2020). Indeed, the five recurrent cytogenetic classes captured using standard-of-care FISH testing predict time-to-first treatment (Hu et al., 2019) and median survival (Döhner et al., 2000), but have not lead to meaningful shifts in clinical use of targeted therapies. One exception to this is the clinical decision to pursue more aggressive chemotherapy for patients with del(17p) and/or *TP53* mutation in advanced disease stages. However, whether this subgroup represents a greater risk in treatment-naïve disease is disputed (Hu et al., 2019).

1.4.2 Current standard of care and theories of therapy failure

Standard care for CLL patients consists of continued monitoring of asymptomatic patients and chemoimmunotherapy for patients presenting with active or symptomatic disease. Prior to 2016, chemoimmunotherapy (fludarabine, cyclophosphamide, and rituximab; FCR) was the standard-of-care for active or symptomatic CLL (Woyach, 2019). However, in the late 2010's, multiple lymphoid-targeted therapies were approved by the FDA, specifically: ibrutinib, venetoclax, and idelalisib. Ibrutinib (originally called PCI-32765) is a chemical inhibitor of Bruton's tyrosine kinase (BTK), a receptor important in B cell development (Honigberg et al., 2010). Ibrutinib has shown promising efficacy in multiple clinical trials as

a monotherapy and in combination (Fraser et al., 2018). However, ibrutinib is associated with cardiotoxicity and severe bleeding in half of patients since BTK is also involved in the coagulation pathway (Brown, 2018). Venetoclax is a chemical inhibitor of B cell leukemia 2 (BCL2), and has also shown promising efficacy as a monotherapy and in combination in phase I clinical trials of patients with relapse/refractory CLL (Fischer et al., 2019; Roberts et al., 2016). However, venetoclax treatment consistently leads to severe neutropenia (abnormally low neutrophil counts) in half of patients (Fischer et al., 2019; Roberts et al., 2016), suggesting multiple cells types are being affected. While these new therapies are improving patient outcomes, treatment is given indefinitely or until toxicity forces discontinuation (Brown, 2018), and complete remission is uncommon (Roberts et al., 2016).

Although standard-of-care therapies often control disease progression, CLL is remains incurable and patients may succumb to immune-related complications or eventual transformation of the disease to an accelerated phase (Rossi et al., 2018). Furthermore, these treatments focus on symptoms of CLL, they do not disrupt the underlying genetic mutation-based molecular targets in leukemic B cells. Selective genetic mutation-based molecular targeting has become a successful standard-of-care in leukemias such as chronic myelogenous leukemia (*BCR-ABL1* fusion protein targeted by imatinib) (Druker et al., 1996) and acute promyelocytic leukemia (*PML-RARA* fusion protein targeted by all-trans retinoic acid) (Wang and Chen, 2008). Similar genetic mutation-based molecular-targeting therapies have not yet been developed for CLL. However, the selective genetic mutation-based targeting of malignant B cells arising from CLL disease would represent a significant advance towards precision therapy. Theories of therapy failure include the long persistence of the disease and therapy-resistant mutations. Persistent active disease is a clinical concern because it is believed to promote the development of drug resistance (Ahn et al., 2017). This is exacerbated through the paucity of early treatment interventions available to CLL patients, which means that acute interventions with intensive immunochemotherapy are commonly the only option. Additional mutations in key genes, such as BTK, negate the effects of targeted therapies such as ibrutinib (Quinquenel et al., 2019). For example, after three years of ibrutinib treatment for CLL patients, only 31% of patient remained on treatment, and of those patients, 57% had acquired mutations in the BTK receptor, which was positively correlated with subsequent disease progression.

1.4.3 Development of experimental models

CLL patient-derived PB MNC samples have been used in a wide array of research studies, as they can be collected from PB and cryopreserved until use (Oppermann et al., 2016; Spaner et al., 2013). Despite this useful resource, drug screening results have yet to identify unified responses common to genetic subtypes, and instead have reported 'strongly diverse' and 'patient-specific' drug responses. This may be attributed to additional genetic modifications that have not been assessed, insufficient power and statistical analysis to distinguish specific subtypes, and confounding responses from non-leukemic residual cell impurities. Indeed, up to 15% of cells assessed in large studies of CLL are not B cells (Beekman et al., 2018), and presumably do not carry the unique genome and epigenome of interest. Despite temporary survival of CLL patient-derived cells in xenografts *in vivo* permitting short term studies (Durig et al., 2007; Ljungars et al., 2018), there is no compelling evidence to suggest the diseased cells can be propagated *in vitr*o or using xenografts. Indeed, cell lines used *in vitro*

in the study of CLL are generally complex karyotype and derived from non-CLL disease such as mantle cell lymphoma, diffuse large cell lymphoma, and acute lymphoblastic leukemias (Honigberg et al., 2010; Souers et al., 2013). Furthermore, human PSC-derived B lymphoid cells lack expression of classical B cell receptors such as CD5 (Carpenter et al., 2011; French et al., 2015), which is a key regulator of the B cell receptor (BCR) pathway, and one of the defining antigens of human CLL.

Experimental models of CLL that can be propagated and as such used more universally have generally been developed using mouse models and transformed cells lines. Mouse models are largely contained to two groups, one of which seeks to produce B cells or B cell-like expansions for chemical targeting, and the other seeks to understand the contribution of small genetic mutations common to human CLL. In the former, multiple transformed cell lines can be successfully transplanted into mice to produce lymphoma tumors which do not carry the genetic mutations unique to CLL, but provide an opportunity to target human B cells *in vivo* (Souers et al., 2013). Mouse models of arthritis have also been used in CLL studies due to the paucity of more matched experimental models (Honigberg et al., 2010). A more widely adopted transgenic mouse model is the Eμ-TCL1 mice (Bichi et al., 2002). In these mice, *T cell leukemia-1* (*TCL1)*, a gene commonly activated by translocations in T cell leukemia, is under the control of an immunoglobulin-based promoter and enhancer region termed 'Eμ'. This targets expression of *TCL1* to immature and mature B cells. Using this model, *TCL1*-driven expansion of CD5+ B cells is observed in adult mice. Lastly, researchers have targeted deletions of sections located within del(11q) or del(13q) into all somatic cells of mice and reported defects in only B lymphoid cells (Klein et al., 2010; Yin et al., 2019). However, the somatic nature of these models deviates from clinical observations.

1.4.4 Therapeutic targeting of molecular drivers

There is great interest in understanding the drivers of CLL from a genetic basis, since genetic models of three of the four main FISH-based abnormal cytogenetic subgroups of CLL have been reported; del(11q) (Yin et al., 2019), del(13q) (Klein et al., 2010), and del(17p) (Zenz et al., 2010). To date, genetic mutation-based targeted therapies have not been developed for these three deletions. Strikingly, no genetic models or genetic mutation-based targeted therapies exist for tri12 CLL. The frequent occurrence of tri12 in CLL, the second most common mutation based on FISH cytogenetic testing, and its exceedingly rare occurrence in other forms of leukemia (Mitelman and Levan, 1981) strongly suggest that genes located on chromosome 12 are involved in CLL. The complex genetic impact and molecular phenotype of tri12 CLL versus relatively small deletions on smaller chromosomes has made modelling of this cancer difficult. Patient samples have provided the only resource for understanding tri12 (Beekman et al., 2018), and are limited by the caveats detailed earlier in this section.

1.5 Chemotherapy-induced myelosuppression

1.5.1 Clinical characteristics and pathophysiology

Chemotherapy and radiation are prominent modes of cancer therapy. However, patients suffer from non-specific off target effects of these cytotoxic therapies that effect heathy cell proliferation required for tissue maintenance (Vahdat et al., 2002). Due to the high cellular turnover of the hematopoietic and immune system, these cells are particularly sensitive to

cytotoxic injury and experience myelosuppression; acute suppression of BM and blood cell production. In fact, chemotherapy-induced myelosuppression (CIM) is the most common dose-limiting side effect of most chemotherapy drugs (Maxwell and Maher, 1992). The prevailing clinical view is that a linear-log relationship exists between high chemotherapy dose and even higher tumor cell kill (Vahdat et al., 2002). This suggests efforts to prevent chemotherapy-induced hematopoietic suppression, and therefore discontinuation of treatment, would enable more aggressive and effective anti-cancer therapy. CIM, and the more extreme state of BM failure, are defined as the inability to produce the blood cells necessary to control immune function (Green and Rubin, 2014). Because of this, infection is one of the most common life-threatening complications of cancer treatment.

1.5.2 Current standard of care

Treatment of CIM is based on managing acute cytopenia; deficiencies in mature blood cells. For example, prophylaxis treatment of antibiotics is given to reduce the chance of infection (Zinzani et al., 2019). To mitigate the effects of anemia, red blood cell transfusions and the cytokine erythropoietin may be given (Green and Rubin, 2014). Third, prophylaxis injections of synthetic G-CSF are commonly given in clinical trials to promote neutrophil recovery and hopefully mitigate grade 3 and 4 neutropenia adverse events (Choi et al., 2014). However, prophylaxis treatments may unnecessarily increase treatment and they have not reliably reduced hospitalization costs nor provided consistent clinical impact (Ozer et al., 2000).

1.5.3 Development of experimental models

While these approaches seek to address the acute symptoms of CIM, the role of cytotoxic damage to HSC is largely unknown but recurrently speculated to lead to HSC exhaustion (Velardi et al., 2018). This question has been technically difficult to approach, as it is that well established that HSCs are already heterogeneous in their degree of self-renewal, life span, and differentiation capacity (Birbrair and Frenette, 2016). Our lab coined the term "hematopoietic regenerating cell" (HRC) to describe the residual cells remaining after chemotherapy, in human-mouse xenografts, that presumably give rise to emergency state hematopoiesis (Boyd et al., 2018). Chemotherapy-treated HRC rebound quite slowly and do not exceed their initial growth rate nor expand above original BM reconstitution level, and was a consistent feature across adult and neonatal sources of HRCs. Preliminary bulk transcriptional profiling of HRC and therapy-naïve populations of healthy cells showed that HRCs differ from therapy-naïve HSCs studied to date, as well as leukemic regenerating cells, highlighting the possibility that there are specific and targetable pathways that may promote healthy regeneration in the absence of leukemic progression. However, these pathways and the emergence of critical subpopulations of HRC have yet to be explored. However, transcriptional profiling of individual non-hematopoietic BM niche cells in mice, in times of homeostasis and following CIM, have provided molecular insight into pathways involved in injury response (Tikhonova et al., 2019). In fact, alternative approaches have sought to repair damaged or affected BM niche cells to indirectly support healthy hematopoiesis, including adipocytes (Boyd et al., 2017), and the nervous system (Lucas et al., 2013).

1.5.4 Theories of targeting hematopoietic stem cells

Bulk hematopoietic recovery has been experimentally improved in xenografted mice using cytokines and growth factors, such as provision of G-CSF or follicle stimulating hormone (Bujko et al., 2019; Maxwell and Maher, 1992) or suppression of luteinizing hormone (Velardi et al., 2018). However, protection of HSCs from chemotherapy is a relatively new concept, and has been attempted using cell cycle inhibition (Brenet et al., 2013; He et al., 2017). Brenet *et al*., used chemical inhibition of transforming growth factor beta, which lead to downstream targets promoting cell cycle inhibition (Brenet et al., 2013). He *et al*., directly used chemical inhibitors of cyclin-dependent kinase 4/6 to transiently suppress HSPC cell cycling while concomitantly administering chemotherapy (He et al., 2017). While this regimen lessened the reduction of BM HSPCs, it is not clear if this would be applicable to cancer patient-derived xenografts treated with common chemotherapies or if it is applicable to human HSPCs as these proof-of-principles were completed using murine HSPCs. Therefore, to date there are no clinically therapeutic options available to prevent or treat the exhaustion of HSPCs as a side effect of CIM and necessary life-saving chemotherapy.

1.6 Bioinformatics

1.6.1 Data mining

The Information Age began in the 1970's where the economy transitioned from being primarily based on industrialization towards today's state of being based on information technology. The biological sciences have followed these trends with the increasing availability of large data, which has created new solutions of how to extract useful knowledge. Simply put, data mining is series of steps taken to manage large data sets, identify patterns, and establish relationships. For example, the modern biological researcher now has access to *The Human Genome Project* and >100,000 sequenced human genomes (Green et al., 2015), *The Cancer Genome Atlas* and >12,000 sequenced tumor samples (Hoadley et al., 2018), and the *Gene Expression Omnibus* with >1.3 million publicly deposited gene expression data sets from over 2000 organisms (Clough and Barrett, 2016). Data mining includes searching for patterns in large data, and not just evaluating automated results. For example, data mining techniques pioneered by Sokolov *et al*. (Sokolov et al., 2016a; Sokolov et al., 2016b), were successfully applied to The Cancer Genome Atlas, which extracted a "stemness index" by which to rank order samples in a process the authors termed "oncogenic dedifferentiation" (Malta et al., 2018). These data sets are too large to manually comb though, and so techniques of data mining are necessary and fundamental to biological research; some of which this thesis introduction will cover in the following subtopics.

1.6.2 Population and single-cell analysis

RNA analysis was once limited to the study of individual transcripts using Northern blotting and quantitative polymerase chain reaction (PCR) (Tachibana, 2015). *Transcriptomics*, the study of all messenger RNA molecules expressed from all the genes of an organism's genome, was not feasible until microarray technology, a high-throughput microchip, could measure thousands of transcripts simultaneously and offered standardized assays. However, two major limitations of microarray technology are that it only reports relative gene expression analysis instead of absolute quantitation, and it is based on a series of known transcripts and so novel sequences (i.e. mutations and single nucleotide polymorphisms, SNP) would not

always be reported. Development of next generation sequencing (NGS), based on the principle of detecting nucleotide bases when DNA polymerase adds them to a strand of cDNA, but executed in a massively parallel flow cell, enabled RNA-sequencing (RNA-seq) at relatively low cost. RNA-seq of bulk cells improved upon the technical limitations of microarray gene expression. Gene expression could now be quantified in terms of number of transcripts per gene as a fraction of all the reads measured and mapped to the genome. Furthermore, complex transcripts which deviate from "normal" could be captured, including alternate splicing, SNPs, and mutations.

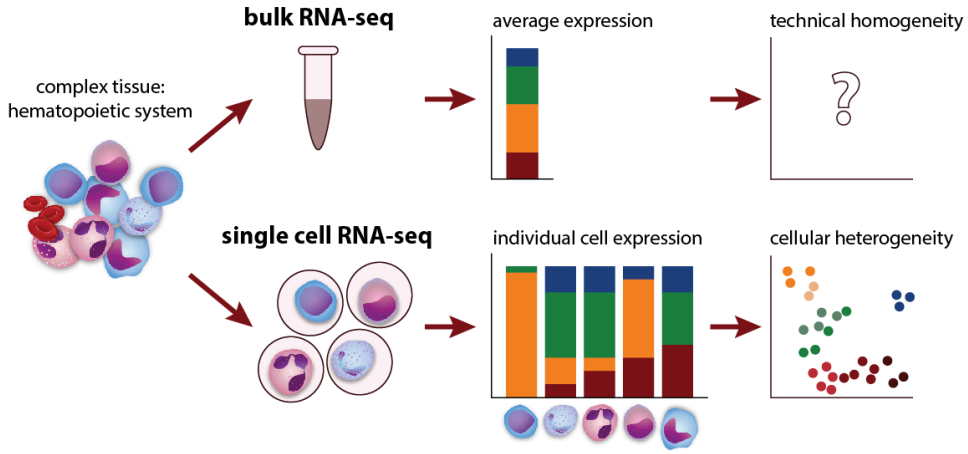


Figure 1.6.2 Contrasting data obtained by bulk and single cell RNA-seq

The immense benefit of single cell RNA-seq (scRNA-seq) is intuitively clear when considering the many intercellular differences across a population of cells, and especially across an entire organ system; multiple cell types, cell cycle differences, responses to stimuli, *etc.* (**Figure 1.6.2**). ScRNA-seq enables the measurement of thousands of transcriptomes simultaneously, and is particularly suitable for identifying potentially useful markers, such as cell surface proteins, for populations of cells with unknown markers. (AlJanahi et al., 2018). Although scRNA-seq is an exciting and novel technology which captures the exquisite complexity of heterogeneous cells, the quality of scRNAseq data is typically much lower than traditional bulk RNA-seq (Tracy et al., 2019). Gene dropout of lowly expressed genes (systematic error) as well as even highly expressed genes (due to sequence-specific reverse transcription difficulties), have not been solved to date. Tracy *et al.*, developed an algorithm which "fills in the gaps" based on closely related cells, a process known as imputation, with the assumption that expression patterns are equally shared across present and absent transcripts in the data set. However, this is not common practice. Therefore, it is important to note that the absence of a transcript in scRNAseq data does not always mean that the transcript is not present. Therefore, follow-up techniques with greater detection limits (such as PCR) and demonstration of protein translation (such as flow cytometry) are important.

1.6.3 Machine learning and statistics

NGS data from one S4 flow cell can sequence 20 billion reads, cover 400 transcriptomes, require 3 terabytes just to store the raw reads, and this process can be completed in less than two days (NovaSeq 6000, Illumina, technical report). With data sets becoming larger and

larger, patterns and relationships in the data are becoming less feasible to observe manually. Machine learning algorithms are an application of artificial intelligence and have achieved great success particularly in genome-wide association studies (GWAS). For example, researchers from Toronto used machine learning to build a predictive classifier based on the SNPs of patients with major depressive disorder, which could predict medications more likely to benefit each patient as opposed to the current practice of "trial and error" (Maciukiewicz et al., 2018).

A key application of machine learning is *regression*. Regression models are used to predict a continuous value, such as a "stemness index" score, based on many variables, such as NGS-based gene expression (Malta et al., 2018). Sokolov *et al*., pioneered one-class logistic regression (OCLR) which could be trained to identify transcriptional traits of minor subclones, perhaps from a tumor biopsy, within samples containing mixtures of cell types (Sokolov et al., 2016b). This work was further expanded with *regularization*, another form of regression. To avoid capturing noise and over fitting data, Sokolov developed a regularization regression model called "the generalized elastic net" which emphasizes the role of mechanistically interlinked genes and signaling pathways as opposed to individual genes expressed independently (Sokolov et al., 2016a). A third example of regression is the removal of data variation deemed as "noise". For example, removing variation such as cell cycle state and mitochondrial genes is common in scRNA-seq analysis (Stuart et al., 2019).

1.6.4 Dimensionality reduction

Dimensionality reduction (DR) is the process of reducing the number of variables of a data set without a significant loss in overall information. DR allows visualization of the data, as well as identifying the leading variables with the most significant influence over the data set. Two established techniques of DR are principal component analysis (PCA) and *t*-distributed stochastic neighbor embedding (*t*-SNE), while uniform manifold approximation and projection (UMAP) superseded *t*-SNE this past year as the dimensionality reduction tool of choice for scRNA-seq data (Becht et al., 2019) (**Figure 1.6.4**).

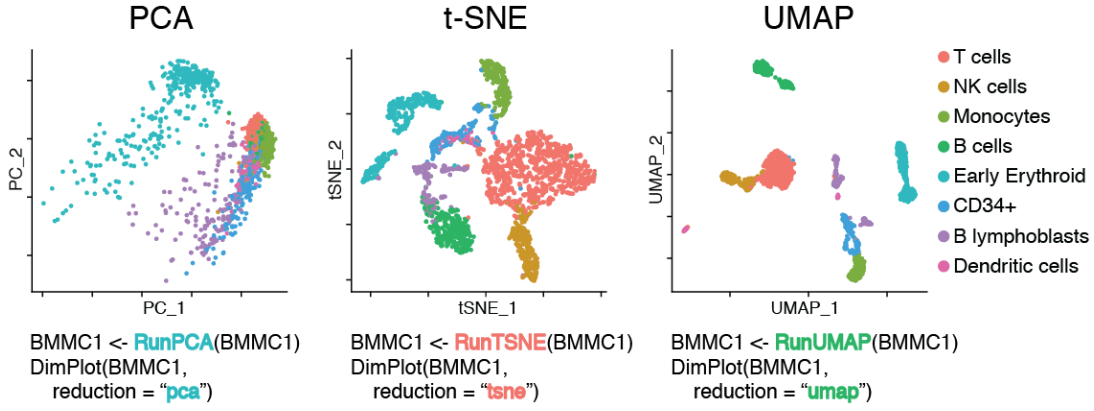


Figure 1.6.4 Contrasting data obtained by dimensionality reduction. The same healthy bone marrow scRNA-seq data represented using three different DR techniques. Raw data from Zheng *et al*. (2017), original analysis by J.C.R.

PCA was developed over a century ago, and is a fast and linear DR technique (Pearson, 1901). Variables are segregated into principal components, ranked by their influence of the

data set, with the first principle component having the most variation, and each ensuing component having the highest variance possible under the constraint that it is orthogonal to the preceding components. This DR technique assumes linear scaling which is unlike biological data such as NGS-based gene expression, where non-linear DR such as *t*-SNE is applicable in highly dimensional space such as scRNA-seq data (van der Maaten and Hinton, 2008). UMAP supersedes *t*-SNE since it is less computationally intensive (i.e. faster), represents global architecture of data better than *t*-SNE, and is built for machine learning based on the clustering of data points instead of *t*-SNE which is mainly a visualization technique (Becht et al., 2019). Overall, the purpose of DR is to highlight the essential characteristics of data to improve understanding and pattern recognition.

1.7 Summary of Intent

The focus of my thesis was to understand components of hematopoietic regulation that are too invasive or complex to study in patients. Canadian research in stem cells, especially HSCs, is strongly engrained in many research programs today, as well providing the basis of many foundational techniques used throughout this thesis; CFU assays (Eaves et al., 1999), SRC assays (Bhatia et al., 1997), and hematopoietic differentiation of human PSCs (Chadwick et al., 2003). In the past twenty years, the ability to capture genetic disease in pluripotent cells as well as regulate self-renewal and manipulate differentiation have built the foundation for my thesis. Despite these milestones, we don't understand the effects of specific genetic mutations in human PSCs, and whether/how they relate to cancer. Furthermore, the field has not been able to produce human PSC-derived HSCs with SRC capacity approaching somatic HSCs. Last, we also lack fundamental insight in the regulation of somatic HSCs in how they overcome injury, which has applications in evolutionarily conserved regeneration as well as identifying key cues that human PSC-derived HPCs (not HSCs) lack. Equipped with this insight, I sought to develop models to understand the regulation of healthy and malignant hematopoiesis. Three central questions provided me with a solid starting basis to characterize key areas of healthy and malignant hematopoiesis:

> *1) Do leukemic mutations captured in the pluripotent state reflect clinically relevant malignant regulation of hematopoiesis?*
> *2) Do proteins which enable regulation of HSC promote the production of human PSC-derived HPCs (or HSCs)?*
> *3) Are there key cells and pathways conserved during hematopoietic recovery from diverse myelosuppressive injuries?*

Shortages of HSCs from healthy donors, chemotherapy-induced myelosuppression, and a lack of genetic mutation-based targeted therapies for leukemia patients, all share a common problem of deficiencies in understanding complex regulation of healthy and malignant hematopoiesis, and reinforce the immediate need for innovative cellular models to address these clinical problems. Understanding whether insights generated in human PSCs can be translated to clinical problems, and vice versa, can be experimentally applied to understand regeneration and healthy blood production. Based on this knowledge, *I therefore hypothesized that human stem cell models provide a focused approach to interrogate the regulation of hematopoiesis from the apex of the hierarchy, which can be used to understand the promotion of healthy hematopoiesis and understand malignant transformation.* If this is

true, then parsing apart these complex issues should identify actionable targets and pathways which would at least incrementally advance our understanding of hematopoiesis *in situ*. To address my hypothesis, I defined the following objectives:

1) Investigate trisomy 12 in human PSCs, with an isogenic counterpart, to identify actionable targets to inhibit malignant trisomy 12 CLL.

2) Apply regulators of healthy HSC-driven hematopoiesis identified to human PSC-derived hematopoiesis to determine sufficiency, improved functionality.

3) Investigate hematopoietic regenerating cells following injury in human-mouse xenografts, to identify regulators of healthy hematopoiesis.

By applying bioinformatic approaches, chemical genomics, and isogenic-controlled experimentation, we demonstrated that trisomy 12 in human PSCs is a robust discovery platform, which enabled unprecedented elucidation of cell autonomous pathways for precision oncology, specifically in the context of tri12 CLL (*Chapter 2*). Together with traditional gene expression analysis, we also developed machine learning algorithms to molecularly dissect pathways specific to tri12 in PSC and CLL, that could only be found using human isogenically-controlled PSC with a genetic leukemic driver *in vitro*. Two actionable targets were therapeutically applied to target tri12 CLL patient samples and tri12 PSCs both *in vitro* and upon xenotransplantation as tumors into immunocompromised mice.

Following this body of work, we then investigated what is necessary to improve PSC-derived hematopoietic progenitors to extend their application beyond *in vitro* models, by applying known regulators of healthy HSC-driven hematopoiesis (*Chapter 3*). We discovered that PSC-derived progenitors fail to survive even 24 hours following bone marrow transplantation, while they can survive and proliferate for weeks *in vitro*. We then determined PSC-derived progenitors do not have a key receptor that healthy progenitors have, called CXCR4. By correcting CXCR4 signaling, we observed enhanced PSC-derived progenitor function and survival in the bone marrow of human-mouse xenografts and a transcriptional shift toward human progenitor profiles. Our results reveal CXCR4 is an actionable target for transplantable progenitors with *in vivo* function.

To investigate hematopoietic regulation in the context of myelosuppression, a pervasive side-effect of radiation and chemotherapy, we developed a third human cellular model using long-term human-mouse xenografts given these broad effects cannot be recapitulated *in vitro* (*Chapter 4*). We sampled xenografts at multiple time points before and during the recovery process after treatments and observed drastic yet reproducible changes in the composition of human progenitors and also mature human hematopoietic cells, which reflected clinical leukemia patient experiences. Furthermore, using single cell transcriptional and protein analysis, we were able to identify unique subsets of human progenitors which may prove to be the key hematopoietic regenerating cells upon further biological investigation of these actionable targets. Validation of clinically relevant interventions which aid in the process of healthy hematopoietic regeneration *in vivo* are the next step in this work.

Overall, this thesis and the data presented within it offer a deeper conceptualization of human stem cell models and the deconvolution of several complex components of hematopoietic regulation, both in healthy and malignant cells. This work has revealed novel, clinically relevant, and actionable targets to ultimately enable the promotion of healthy hematopoiesis on multiple fronts.

Chapter 2: Identification of molecular targets of trisomy 12 transformation in CLL patients using human pluripotency

Preamble

This chapter is a submitted manuscript. It is presented in its submitted form.

Author contributions: J.C.R., D.G., A.L.B., and L.O. performed experiments. L.A. provided clinical annotation. M.G., L.A., F.M., and A.N. provided patient annotation and clinical expertise. J.H., and B.L. provided patient samples and clinical expertise. J.C.R. and M.B. designed experiments, interpreted data, and wrote the manuscript. M.B. directed the study.

Author contributions in greater detail: J.C.R. wrote the manuscript and performed all experiments except for the following: In 10% of *in vitro* experiments, D.G. prepared media, buffers, and antibody mixes for flow cytometry; cryopreserved cells; and thawed cells. D.G. also harvested spleens from CLL PDX mice while I was measuring freshly collected blood. A.L.B. provided feedback on my statistical analyses, performed gavage on my tumor mice for 4 of the 21 days, and provided feedback on the manuscript. L.O. performed the early differentiation screen and provided feedback on the manuscript. L.A. provided clinical annotation for patients #160 and #161, and B.L. provided these samples. M.G., L.A., F.M., and A.N. provided patient annotation and clinical expertise for patients #1 to #159. J.H. provided patient samples and clinical expertise for patients #162-172 and #178-189, and provided feedback on the manuscript. M.B. directed the study and provided feedback on the manuscript.

Identification of molecular targets of trisomy 12 transformation in chronic lymphocytic leukemia patients uncovered using human pluripotency

Authors: Jennifer C. Reid[1,2], Diana Golubeva[1,2], Allison L. Boyd[1], Luca Orlando[1], Andrew Leber[1], Josée Hébert[3,4], Massimo Gentile[5], Luca Agnelli[6,7], Fortunato Morabito[5,8], Antonino Neri[6,7], Brian Leber[9], Mickie Bhatia[1,2*]

Affiliations:

[1]Stem Cell and Cancer Research Institute, McMaster University, Hamilton, Ontario, Canada.

[2]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada.

[3]Department of Medicine, Université de Montréal, Montreal, Quebec, Canada.

[4]Division of Hematology-Oncology, Maisonneuve-Rosemont Hospital, Montreal, Quebec, Canada.

[5]Biotechnology Research Units, Aprigliano, Cosenza, Italy.

[6]Department of Oncology and Hemato-oncology, University of Milan, Milan, Italy.

[7]Hematology Unit, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy.

[8]Hematology and Bone Marrow Transplant Unit, Augusta Victoria Hospital, Jerusalem, Israel.

[9]Department of Medicine, McMaster University, Juravinski Hospital, Hamilton, Ontario, Canada.

*Corresponding author.

Abstract

Genetic aberration-based molecular targeting has become a successful standard-of-care in human leukemias. This is best exemplified clinically in chronic myelogenous leukemia (*BCR-ABL1*) and acute promyelocytic leukemia (*PML-RARA*). However, in the most common adult hematologic malignancy, chronic lymphocytic leukemia (CLL), patients have not benefited from similar genetics-based targeting therapies to date. While lymphoid cell-targeted therapies have improved disease management, druggable underlying genetics-based targets have yet to be identified, including those for somatic trisomy 12 (tri12) found in a quarter of CLL patients. Here, we utilized a unique platform with the tri12 cytogenetic aberration in human pluripotent stem cells (hPSC) and validated targets in this subset of CLL patients. From a cohort of 159 patients, tri12 specific autonomous signaling was shown to strongly influence clinical outcomes with a signature that could be uniquely captured in isogenically controlled tri12 hPSCs. Using tri12 hPSCs as a quantitative framework, we identified clinically actionable targets for successful inhibition of leukemic cells specific to tri12 CLL, and independently validated in an additional 30 patients. Selectivity treatment was functionally tested using tri12 and isogenic hPSC co-transplantation in xenograft models, and anti-leukemic efficacy tested *in vivo* using CLL patient derived xenografts. Our study establishes a paradigm using hPSCs to uncover targets from complex genetic networks of human cancers such as tri12 CLL, that have translational potential as biomarkers as well as therapeutic drug intervention.

One Sentence Summary

Selective genetic aberration-based therapeutic targets for genetic group of chronic lymphocytic leukemia patients can be uniquely revealed using isogenically controlled human pluripotent cells that contain the same genomic aberration.

Introduction

Despite continuing development of knowledge from the study of human cancer 'omics' (1), identifying a link between genetic features of individual cancers with specific treatment options remains the single largest challenge (2). Chronic lymphocytic leukemia (CLL) is the most common adult hematologic malignancy, where leukemic B cells accumulate and overgrow healthy cells. This imbalance of dysfunctional B cells leads to progressive bone marrow failure and acquired immune dysfunction (3), with a subset of patients undergoing transformation toward an accelerated terminal phase of disease (4). Unfortunately, current standard of care for CLL patients is limited to monitoring asymptomatic patients and only providing treatment for patients presenting with active or symptomatic disease. Recently however, the therapeutic algorithm for CLL began shifting (5). Chemoimmunotherapy (fludarabine, cyclophosphamide, and rituximab) remains an option for low-risk patients while new drugs including BCL2, BTK, and PI3Kδ inhibitors improve disease management and add to the therapeutic armamentarium, however these are not CLL patient specific (6-9). Challenges remain using these pan-inhibitors due to cardiac and hepatic toxicity, immune dysfunction, opportunistic infection(s), partial responses, and drug resistance, leading to therapy discontinuation, and leaving CLL incurable for most patients (7, 10-16). Traditional

chemotherapy, as well as novel targeted therapies focus on the symptoms of CLL, but do not inhibit underlying genetic aberration-based molecular targets in leukemic B cells (12). As in other leukemias such as chronic myelogenous leukemia (CML: *BCR-ABL1*) (17) and acute promyelocytic leukemia (APL: *PML-RARA*) (18), selective genetic aberration-based molecular targeting of CLL disease would represent a significant advance. However, unlike CML and APL, CLL patients lack a common genetic aberration.

CLL is segregated by four common molecular lesions: trisomy 12 (tri12), del(11)q, del(13)q, and del(17)p assessed by fluorescence *in situ* hybridization (FISH) as part of standard of care; the 5th subtype is devoid of these aberrations ($FISH^{NEG}$) (19). Although these lesions are not therapeutically targeted, they provide prognostic biomarkers incorporated into decisions for standard of care, which includes detection of mutation status of immunoglobulin heavy-chain variable region (*IGHV*) (20). With and without mutated *IGHV*, tri12 in CLL represents a common genetic abnormality present in a quarter of CLL patients (19) typically as the sole genetic lesion (21), and is also common in a preleukemic state of CLL termed monoclonal B lymphopoiesis (MBL) (22). Genomic tri12 is otherwise embryonically lethal in humans (23), with low grade tri12 mosaicism at birth (24), and tri12 non-Hodgkin's lymphoma (25), being extremely rare and only reported in patient case studies. Accordingly, human somatic tri12 is unique to CLL disease pathology (26). Despite several efforts to identify specific features of tri12 in CLL (1, 21, 27, 28), there are no candidate targeted therapies, and these CLL patients continue to experience shorter time-to-first treatment (29) and intermediate prognosis (19). Several model systems using immortalized non-CLL lymphoma cell lines, non-CLL mouse models (arthritis and lupus), a dog model of lymphoma, and/or CLL patient samples treated *in vitro* (Table S1), three leading first-in-class lymphoid-targeted therapies (venetoclax, BCL2 (30); ibrutinib, BTK (31); and idelalisib, PI3Kδ (32)) were developed, but all three present with clinical challenges (13, 33, 34). This outcome may be due to the indirect nature of the preclinical therapeutic potential for CLL patients used in these models that lack surrogacy required to achieve specific drug efficacy. Therefore, continued efforts to apply precision oncology are hampered in part from limited preclinical CLL models that allow targeting of genetic drivers in CLL. Instead, current lymphoid-targeted therapies treat the palliative symptoms of CLL.

Patient-derived xenografts (PDX) of human cancer cells transplanted into immunocompromised mouse models have become a gold standard for demonstrating preclinical therapeutic potential in several hematological malignancies (35), with the conspicuous exception of CLL (36). Unlike PDX models, which are not high throughput discovery tools, *in vitro* drug screening with CLL patient samples expands throughput, however, results in perceived 'patient-specific' responses (37) underscoring the need for new models. Human (h)PSCs offer the ability to study the effects of a cytogenetic aberration, while holding constant any effects in the underlying genetic background; a technique which is not common in CLL disease research. Human PSCs with tri12 afforded the ability to overcome previous barriers in studying tri12-specific transformation in CLL by using an isogenically controlled hPSC line as the basis for a therapeutic discovery platform. Combined with unbiased machine learning, transcriptomics, chemical genomics approaches, *in vivo* drug testing of CLL patient-derived xenograft mouse models, *in vivo* drug testing of tri12 and isogenic PSC-derived tumors in co-transplant xenograft mouse models, and over 180 deeply characterized CLL patient samples have allowed us to delineate tri12 cell autonomy in

shorter progression-free survival, malignant lymphopoiesis, and identify tri12-selective targets for the first time.

Results

Isogenically controlled tri12 gene networks underlie disease progression in patients

Reliable statistical models for time-to-progression and time-to-first-treatment in CLL are of high clinical interest but are prone to variations (38). Although expression of unique molecular targets from diseased tissue have successfully predicted progression-free survival (PFS) in other cancers (39), this has not been the case for CLL. Initially reported at 3.8 years (40), we accrued 159 patients MBL and CLL patients where the cytogenetics and transcriptome of highly purified peripheral blood (PB) leukemic B cells were profiled upon diagnosis, and patient outcome was monitored for up to ten years (Table S2). Approximately half of CLL patients are classified as *IGHV*-unmutated, which is not an oncogenic driver, but patients experience significantly worse prognosis than *IGH*V-mutated patients (3). Further analysis revealed tri12 to be the most common genetic abnormality in the *IGHV*-unmutated cohort, and these patients experience significantly worse PFS (Fig. 1A). As there are no current validated gene expression signatures which predict PFS in tri12 CLL patients; these patients have remained indistinguishable from the absence of tri12 (Fig. 1B), consistent with the intermediate prognosis of these two subtypes (19). We next performed hierarchical clustering to determine whether disease progression could define a subpopulation of these patients based on the transcriptome of highly purified B cells from a cohort of 159 patients (Fig. 1C). However, no trends were discernible, nor were we able to discern subpopulations of these patients within the over-represented subtypes of tri12 and $FISH^{NEG}$ using principal component analysis (Fig. 1D). This suggests a new approach is required to define common molecular traits shared between patients with the same genetic aberrations; i.e. tri12.

Beyond CLL, acquisition of somatic tri12 is extremely rare in humans (23-25). We therefore evaluated the potential use of a pluripotent model that allows for comparative isogenic control cells that are absent for tri12 abnormality. This allows identification of gene expression patterns specific to tri12 by removal of background noise found in isogenic controls. Spontaneous acquisitions of genetic aberrations have been documented in hPSCs (41-43), and we identified a genetic variant of hPSCs (44) that acquired non-mosaic (all cells) tri12 with no other chromosomal aberrations (Fig S1, A and B). This variant hPSC was termed 'tri12 hPSCs'. Using parental hPSCs devoid of tri12 that maintained a normal karyotype for over 60 weeks (>1.25 years), an isogenic comparative control to tri12 hPSCs was available and termed 'isogenic hPSCs' (Fig. S1C). Deeper analysis using genomic hybridization, probing for microdeletions and insertions (<25 kb), demonstrated the overall genetic stability of tri12 hPSCs over continuous long-term culture for >8 months (Fig. S1D). As isogenic controls are unavailable for the study of leukemic cells from CLL patients, we envisioned tri12 vs. isogenic hPSCs provided an unprecedented high-resolution platform to understand the specific genetic contribution and differences in transcriptional networks governed by tri12 alone.

As hPSC variants have been reported to acquire features of transformation (41, 42), we measured these properties of neoplastic hPSCs in tri12 hPSCs. In contrast to transformed hPSCs (41, 42), tri12 and isogenic hPSCs share typical normal hPSC colony morphology (Fig.

1E), and were similar during routine passage as measured by a battery of biological assays including proliferation (Fig. S1, E and F), expression of pluripotent hallmarks (Fig. S1G), or resistance to apoptosis (Fig. S1H). To initially determine whether tri12 induces changes in the pluripotent state that could biologically impact tri12 hPSC behavior, we used multiple differentiation-inducing conditions to evaluate germ layer lineage potential: FoxA2 (endoderm), Pax6 (ectoderm), and Runx1 (mesoderm; Fig. 1F) (45). Tri12 hPSCs readily differentiated as shown by efficiently acquired lineage markers, most notably towards mesoderm (Fig. 1G), as well as down-regulating pluripotent factor, Oct4 (Fig. S1, I and J). In vivo, teratoma-based differentiation was tested by injecting hPSCs subcutaneously into the flank of immunocompromised mice. We observed tissues of all three germ layers represented, within the same tumor, of both tri12 and isogenic hPSC-derived teratomas (Fig. S1K). This suggests that tri12 aberration does not lead to overt malignant transformation (42), but impacts biological networks involved in mesodermal development alone.

To investigate the complex effects of a large-scale genetic lesion such as tri12 seen in CLL, but in the absence of patient heterogeneity, we calculated differentially expressed genes (DEG) that were up-regulated in tri12 hPSCs, compared to uniquely available isogenic hPSCs lacking tri12 (Fig. 1H). This allowed us to define an isogenically controlled tri12 gene list (Table S3). Using these highly refined tri12-specific genes, we were able to segregate tri12 from $FISH^{NEG}$ in leukemic B cells purified from CLL and MBL patients (Fig. 1I). This segregation could not be accomplished solely using global transcriptional of leukemic B cells (Fig. 1D), consistent with approaches used in previous studies (1, 28, 40). These data and applied patient validation raised the opportunity that commonly activated genes in tri12 CLL patients may be shared in tri12 hPSCs and used for investigation.

Tri12 gene dosage in PSC enables a classifier for CLL patient disease

To identify whether there are underlying and shared perturbed gene networks reflected in tri12 CLL using isogenically controlled tri12-specific genes from hPSCs (Table S3), we performed gene set expression analysis (GSEA). As GSEA determines whether statistically significant concordant differences between two biological states may exist based on a defined set of genes (Table S3), we extracted tri12-only and cytogenetically normal $FISH^{NEG}$ patients from our larger cohort (Table S2). GSEA demonstrated genes in the hPSC-based tri12 gene list to be highly enriched in tri12 CLL compared to $FISH^{NEG}$ CLL patients (Fig. 2A) and enriched in tri12 MBL compared to $FISH^{NEG}$ MBL patients (Fig. 2B). These analyses indicate that key genes perturbed by tri12 aberration are shared between hPSCs and CLL patients and can be identified based on the use of the isogenically controlled hPSC system.

To understand the relationship of tri12-specific gene expression and patient outcome, genes common to tri12 CLL and MBL patients from the GSEA leading edge (LE) were evaluated within the cohort of patients with leukemic cells with tri12 or $FISH^{NEG}$ genetics, previously shown to have equivalent prognosis (Fig. 1B). While the LE genes individually trended towards higher expression for tri12 patients who also experienced disease progression (Fig. S2), the sum of these genes could significantly identify tri12 patients at risk of disease progression (Fig. 2C). PFS was assessed by stratifying patients solely based on LE gene expression and observed significantly shorter PFS when LE genes were expressed (Fig. 2D). The differences in survival were significantly attributable to LE gene expression using multivariate Cox proportional hazards analysis (Hazard ratio 2.84, $P<0.025$). This

multivariate analysis also found equivalent risk for other features (Rai stage 0 CLL vs. clinical MBL, and normal vs. tri12 cytogenetics), consistent with previous reports (40, 46). These findings indicate tri12 molecular signatures from hPSCs provide significant clinical segregation and stratification of CLL patients.

To compare all cytogenetic groups of CLL and MBL in an unbiased global-transcriptome approach, we used a machine learning algorithm known as one-class logistic regression (OCLR; Fig. 2E), which recently identified a stemness index for cancer using hPSCs (47). The OCLR machine learning algorithm trained a model based on tri12 hPSCs that incorporated a regularization scheme which emphasizes the role of mechanistically interlinked genes (48) to develop a Tri12 Score (Table S4), ranging from low (zero) to high (one). We validated the Tri12 Score by applying it to our clinical dataset composed of leukemic cells from the 159 individual MBL and CLL patients (40), and by comparing the genetic subtypes separately. Importantly, higher values for the Tri12 Score were only associated with biological samples from patients with leukemic cells harboring tri12 (Fig. 2, F and G). Other CLL and MBL genetic subtypes scored lower, further supporting the uniqueness of the tri12 CLL transcriptome (1), but also in clinical MBL cases of tri12. Importantly, we determined that tri12 score was not an artefact of *IGHV* mutational status (Fig. 2H), or associated with the presence of CD38, ZAP-70, *NOTCH1* mutations, or *SF3B1* mutations previously known (Fig. S2B, Table S2). These observed distinctions are especially important due to our previously reported effects of *IGHV* mutational status on patient transcriptional profiles (40). Based on this patient clinical validation, isogenically controlled tri12 hPSCs provide a platform to extract leukemic features of cell identity-independent biological processes that are unique to patient-specific leukemic drivers.

Cell autonomous role of tri12 in features of leukemogenesis

Capture of tri12 enables the biological and regulatory dissection of cell autonomous features of leukemic disease *in vitro*. Isogenically controlled tri12 hPSCs exhibit enhanced proliferation, consistent with clonal tracking results in CLL in patients (49) and transcriptional hallmarks of leukemic gene regulation (Fig. 3, A-C). CLL patient cells share a well-defined immunophenotypic profile based upon CD5, CD19, CD20, CD23, and CD79b expression, as well as kappa or lambda (κ/λ) surface immunoglobulin light chain restriction, which identify a clonal population of mature B lymphocytes (3). The cell autonomous impact of tri12 in hPSCs enabled the derivation of B cells, which also expressed hallmarks of malignant lymphopoiesis (Fig. 3, D to L; Fig. S3). Malignant lymphopoiesis was not observed using isogenic hPSCs (Fig. 3) and is consistent with other reports of normal B-lymphoid derivation using cytogenetically normal hPSC (50). The interrogation of tri12 as a leukemic driver in the pluripotent state provides features of CLL patient cellular pathology in a cell autonomous context. These results provide the foundation to investigate whether tri12 and isogenic hPSCs could resolve elusive targets that could be applied to tri12 CLL disease.

Identification of selective chemical targeting of tri12 gene dosage

Despite representing a potential adjuvant for therapy in CLL patients, tri12-specific chemical inhibitors have yet to be developed for CLL disease treatment. To determine whether features of tri12 could be targeted using pathways that were cell identity-independent, we

designed and optimized a chemical screening platform for both isogenically controlled tri12 hPSCs as well as relapse/refractory sex- and age-matched CLL patient B cell samples with tri12 and normal cytogenetics (Fig. 4A, Table 1, Fig. S4). Our goal was to identify highly potent chemical inhibitors of gene targets from isogenically controlled tri12 hPSCs as well as the recently reported reference epigenome of CLL (1) (Tables S2 and S5). A total of 13 compounds were tested for tri12 selectivity based on the availability of known inhibitors identified using The Drug Gene Interaction Database (www.dgidb.org) based on these gene lists (Tables S2 and S5). Lead candidates from chemical screening were inhibitors of endothelin receptor type B (EDNRB), interleukin-1 receptor-associated kinase 4 (IRAK4), and Kit, and validated using CLL patient cells (Fig. 4, B to G; Table S6 and S7). These three receptors were also preferentially expressed at the single cell protein level in hPSCs (Fig. 4, H and I), which was consistent with RNA expression and drug sensitivity (Fig. 4, B to G; Table S2). Mechanistically, neither EDNRB nor IRAK4 have been linked to tri12 CLL. However, elevated endothelin ligand has been observed to be elevated in CLL patient blood, although it is not clear which genetic subtypes of patient CLLs were used in this study (51). The endothelin receptor pathway is involved in cancerous properties such as cell proliferation, invasion, metastasis, and angiogenesis (52). Additionally, the IRAK4 receptor pathway mediates downstream targets in the toll-like receptor and NF-κB pathways, which are known to mediate cancer initiation and progression, and regulates cellular processes such as inflammation, apoptosis, and differentiation (53). We propose these therapeutic targets affect tri12 CLL and PSC biology through gene expression networks enriched in genes located on chromosome 12 (Fig. 4J) and are especially of interest due to their effective chemical targeting and tri12 selectivity, the first of its kind to be demonstrated *in vitro*. This led us to evaluate the effectiveness of targeting human tri12 cells transplanted into xenograft mice to better estimate potential preclinical efficacy.

In vivo chemical targeting of tri12 tumor cells

Patient-derived xenografts (PDX) of human cells transplanted into immunocompromised mouse models have become a gold standard for demonstrating preclinical therapeutic potential in several hematological malignancies (35), with the conspicuous exception of CLL (36). CLL PDXs have permitted short term study of drug interactions in physiological microenvironments (54), however less than 1% of transplanted cells can be recovered beyond 4 weeks *in vivo* (55). While hematopoietic stem cells purified from CLL patients can sustain grafts beyond 4 weeks, these grafts are consistently devoid of the genetic aberrations observed in the donor patients (56). Based on these observations, we designed *in vivo* drug treatment regimens for our lead chemical inhibitors of EDNRB and IRAK4, EDNRB-i (bosentan) and IRAK4-i (PF06650833), respectively.

To first examine the tolerability and selectivity of drug treatment *in vivo*, we first performed teratoma xenograft studies in mice using isogenic and tri12 PSCs transplanted into the same mouse, but located distally from each other within the same treated recipient. Since palpable tumors are robustly formed upon subcutaneous injection of PSCs into the flanks of immunocompromised mice (Fig. S5A), as opposed to other routes of injection (57), we chose this approach (Fig. 5A). PSC-derived tumors were established for 4 weeks and confirmed through palpation prior to the administration of drug treatment, which was administered daily by oral gavage for 21 days at 15 mg/kg for both EDNRB-i and IRAK4-i.

This dose was selected as a low-to-moderate dose based on extended treatment in previous animal studies (58, 59) and phase II clinical trial of arthritis patients (NCT02996500, ClinicalTrials.gov). This treatment regimen was sublethal and did not interfere with mouse health or body weight (Fig. 5B). Upon tumor harvest after 7 weeks, we observed significant pairing of tumor sizes across mice, which was consistent in all treatment groups (Fig. S5B). By co-transplanting both tri12 and isogenic PSCs into the same mouse (but in separate locations), we could normalize inter-mouse variability and assess chemical inhibitor selectivity using an *in vivo* drug delivery system. Therefore, we computed the ratio of tri12 / isogenic tumor weight (Fig. 5C). This resulted in a significant reduction in tri12-specific tumour burden for IRAK-i (*P=0.02). While the majority of tri12 tumors were equally affected between EDNRB-i and IRAK4-i, no significant reduction for the EDNRB-i group was detected.

Expanding on these observations, we employed a human CLL PDX model; rarely used or reported in pre-clinical research due to difficulty of sustain a stable CLL PDX graft. CLL PDX models were established from patient CLL#160, whose leukemic cells harbour tri12, as this patient had persistent active disease for 8 years but had not received any treatment in the 3 months prior to sample donation (Fig. 5D). 100x106 PB MNCs were transplanted intravenously into preconditioned NSG immunocompromised mice and allowed to home to lymphoid organs for 4 days prior to initiation of chemical treatment (54). EDNRB-i and IRAK4-i were administered daily by oral gavage (100 mg/kg p.o.) (60, 61) for 4 days and mice were harvested on the 8th day (Fig. 5D). This transplant format was previously developed by Ljungars et al., (54), and enables the chemical treatment of PDX mice harboring human tri12 CLL. Recent clinical trials for CLL patients have reported common adverse events of reduced platelet counts and anemia (62, 63). Accordingly, we assessed healthy blood parameters and absolute quantitation of circulating white blood cells (WBC) in PB of the PDX mice with a blood analyzer at experimental endpoint. These treatment regimens at the higher dose of 100 mg/kg did not affect hematocrit, red blood cell counts, or platelet counts (Fig. 5F). Significant reduction of CLL in the PB was detected upon treatment with EDNRB-i (**P=0.0006) or IRAK-i (*P=0.0206; Fig. 5G), similar to selective tri12 tumor effects using hPSC tumors (Fig 5E). Finally, we assessed chimerism of the human CLL PDX mice spleens using flow cytometry (Fig. 5H). Both chemical inhibitors separately achieved a significant reduction of CLL engraftment in the spleen (Fig. 5I); EDNRB-i (**P=0.0057) and IRAK4-i (**P=0.0031). These *in vivo* human tumor and PDX models permitted the study of selectivity and efficacy, and collectively demonstrate EDNRB-i and IRAK-i are bioavailable and sufficiently potent suppressors of tri12 CLL disease, providing the first demonstrated *in vivo* efficacy in inhibiting targets biologically linked to tri12 CLL.

Selective targeting of tri12 CLL patient leukemic cells

To robustly test and determine therapeutic potential of these targets in a broader cohort of CLL patient disease, we evaluated their effects using another preclinical gold standard for translation of CLL drugs to date; patient-derived PB leukemic samples (Table S1) (32). We accrued an additional cohort of 30 CLL patients spanning 3 independent clinical sites, and care, and management (Table 1). Three PB samples of B cells from healthy adult volunteers were used as non-diseased controls. Single cell protein level expression of these targets was found to be completely absent in B cells from healthy volunteers, but present in CLL patient

samples (Fig. 6A) suggestive of selective targeting to leukemic cells alone. To evaluate preclinical efficacy and response (Table S1) (32), all CLL patient samples were evenly randomized by standard clinical criteria (Fig. 6B and C; Table 1). Moving beyond FISH detection of the main genetic drivers of CLL, the clinical standard-of-care cytogenetic test, additional whole genome genetic screening of 5 randomly selected $FISH^{NEG}$ CLL patients ensured no genetic abnormalities were present across the patients' entire genome (Fig. 6C, Table 1). We then examined each of the 30 leukemic samples for protein expression of EDRNB and IRAK4, and found elevated expression of both targets in tri12 CLL patient B cell samples (****P<0.0001; Fig. 6, D to G). Furthermore, ten-point IC50 drug responses to EDNRB-i and IRAK4-i were tested for all 30 CLL patient samples, which determined high selectivity indexes of 7.1 and 5.1, respectively; where tri12 CLL was selectively targeted compared to $FISH^{NEG}$ CLL patient samples (Fig. 6, H and I). Linear regression comparing both drug responses on a per patient basis demonstrated significant correlation and consistency (****P<0.0001, Pearson correlation r=0.90, Fig. 6J).

Our results demonstrate the ability for molecular dissection of leukemic drivers in isogenically controlled hPSC that leads to specific identification of unique targets for therapeutic development in tri12 CLL patients for the first time.

Discussion

While tri12 CLL has previously been revealed to possess unique properties characteristic of mutated cells (1, 21, 27, 28, 64-67), targeted therapies have not arisen. As tri12 is embryonically lethal, the molecular networks effected have remained enigmatic in humans (23), leaving the understanding and therapeutic applications for MBL and CLL patients with recurrent somatic tri12 challenging (19, 22). Despite the ability to use patient derived cells, the complexity of this genetic aberration, and background genetics per patient has made its mechanism of action and role of tri12 in CLL disease unclear. Small monogenic oncogenes captured in the pluripotent state have enabled investigation of disease (68) but capturing tri12 specific affected networks beyond primary CLL patient cells has proved difficult, and biologically restricted for human study. Our current approach demonstrates the ability to capture important aspects of a complex cytogenetic aberrations observed in human cancers, such as tri12, that can be pragmatically applied to uncover cell autonomous features of cancer, as well as provide targets for chemical inhibition and biomarkers of disease progression. Since lead inhibitors for targets identified in our study are preclinical and FDA approved drugs for non-CLL diseases (69), e.g. IRAK4-i in phase II trial for arthritis (NCT02996500, ClinicalTrials.gov), this level of drug cross-label repurposing holds more immediate potential for phase I testing in tri12 CLL patients (Table S7).

The two targets identified here provide proof-of-principle for CLL patient applications and reveals the importance of using isogenically controlled genetic systems to find selective compounds and novel targets. Tri12 and isogenic hPSCs offer the potential of high-throughput automated screening platforms for pre-clinical testing and validation studies for new drugs and combinations moving forward. Co-transplantation of tri12 and isogenic hPSCs for tumor studies in the same recipient mice provide establishment of an accompanying *in vivo* assay for broader testing and application of treatment for precision oncology that assure selectivity to transformed cells. For CLL, while genetic lesions are undeniably strongly correlated with patient outcomes (19), intronic polyadenylation (70) and microenvironment

(71) also affect healthy and leukemic B cell regulation and survival. These factors, in addition to treatment history and patient background, contribute to the difficulty in molecular target identification using CLL patient-derived cells, and further support the rationale of effectively targeting cell autonomous pathways, which has not been accomplished to date in the absence of isogenically controlled systems. Evolution of these targeted therapies will necessitate the need to move beyond genomics in order to determine all the contributors of CLL pathogenesis. Our study using hPSCs establishes a foundation to further validate additional putative targets in biological assays that can be tested and applied to CLL patients. We propose that tri12-selective targets will play an important role in optimized combination therapy specificity; to spare healthy immune cells and reduce side effects currently experienced by patients. This approach also offers a plausible basis to screen for targets in the preleukemic state, allowing for early intervention of MBL patients at high clinical risk for leukemic progression (22).

We propose that tri12-selective targets will play an important role in at least three major areas of precision medicine: (i) better disease management using tri12 biomarkers, (ii) therapeutically targeting tri12 gene targets in both CLL and MBL, and (iii) to adjuvant and synergize with combination therapy in tri12 cases of disease progression and relapse. Identification of novel antigens that result from mutations and cytogenetic aberrations represents an appealing clinical strategy for combination therapy (72), since the expression of antigens such as CD20 are suppressed by the BTK inhibitor ibrutinib (73), and anti-CD20 drugs are not clinically effective when combined with ibrutinib (74). Our findings establish a paradigm for using hPSCs and related genetic techniques to isolate contributions of cancer-associated mutations and cytogenetic aberrations that can be compared and controlled by isogenic counterparts, ultimately allowing translation to other genetically driven cancers with unknown disease related mechanisms.

Materials and Methods

Study Design

The objective of this study was to investigate the effect of tri12 on cell autonomous signaling in human PSC and CLL and identify druggable targets; IRAK4 and EDNRB. Three models were used to address whether IRAK4 and EDNRB affect tri12 CLL: (i) using transcriptomics of purified B cells from CLL patients, (ii) *in vitro* chemical treatment, and (iii) *in vivo* chemical treatment using PDX mice. *Sample size*: Power of Analysis using two different calculations (Stat UBC and ClinCalc), determined 4 additional patient samples per group (tri12 and normal) would substantiate our drug testing claims; power=0.95 and alpha=0.05. This was determined using the GI50 and variance values of tri12 and normal CLL patient B cells (CLL#1 and CLL#2). In total, we screened N=12 tri12 CLL patient B cell samples, and N=18 $FISH^{NEG}$ CLL patient B cell samples (Table 1). *Rules for stopping data collection*: No experiments were stopped prematurely. *Data inclusion/exclusion criteria*: Patient data and samples were included based on standard diagnostic criteria, independently assessed by clinicians. No patient data were excluded. *Outliers*: No outliers were removed. *Selection of endpoints*: Drug screening was completed in 4 days based on our previous screening platform (45). Mice were killed 7 weeks after injection of hPSCs due to large tumor size, and consistent with our previous study (45). CLL PDX mice were killed 8 days after transplant, consistent with previous study (54). *Replicates*: 'N' experiments represent independent experiments completed at a different time with different reagents, 'n' biological replicates represent testing in separate wells completed concurrently but analyzed separately. *Research subjects or units of investigation*: 'N' number of patient samples tested indicates different people. 'N' number of mice tested indicates different mice. *Experimental design*: controlled laboratory experiments. *Randomization*: Mice were assigned randomly to the various experimental groups, based on weight and sex, and distributed among cages. *Blinding*: Clinical colleagues were blinded to the study purpose and results when providing patient samples for analysis. Animal caretakers were blinded to the allocation of mice. Investigators who assessed, measured, and quantified the results were not blinded to the intervention, with the exception of drug screening analysis which was completed with coded ID's revealed after analysis.

CLL patient samples

Tier 1 BRISQ reporting guidelines were used to describe CLL patient characteristics (Table S2) and biospecimen handling methods. Heparinized peripheral blood (PB) was obtained from consenting CLL patients and healthy volunteers (Juravinksi Hospital and Cancer Centre, Hamilton, Ontario), and protocols were approved by the McMaster Research Ethics Board. Mononuclear cells (MNC) were collected and isolated as previously described (75), cryopreserved in 10% DMSO and stored in liquid nitrogen until analysis (5 years, -150°C). PB MNC from CLL patients were also provided by the Quebec Leukemia Cell Bank as well as the Manitoba Tumor Bank, which were cryopreserved in 10% DMSO and stored in liquid nitrogen (9±6 years). Cells were thawed and enriched for CD19+ cells if <95% CD19+ (Miltenyi, cat# 130-050-301), cultured at $1.0x10^5$/mL in serum-free AIM-V medium with ß-mercaptoethanol in 96-well optical glass plates (Corning) at 37°C in 5% CO2. Drug testing was conducted for 4 days, with drugs and doses indicated in the Figure legends. In a subset of experiments, CLL cells were stimulated with 2S conditions (37); 500 U/mL interleukin-2 (Sigma) and 1 µg/mL imidazoquinoline (R-848; Sigma) and quantified after 2 days.

Human PSC culture

Human embryonic stem cell (hESC) research received Canadian Stem Cell Oversight Committee (SCOC; Canadian Institutes of Health Research, CIHR) approval, and Research Ethics and Biohazard Utilization Protocols approval at McMaster University, following the principles of the 2016 ISSCR Guidelines for Stem Cell Research and Clinical Applications of Stem Cells. CA2 and H9 hESC were cultured in feeder-free conditions in mouse embryonic feeder (MEF)-conditioned media (MEF-CM) with 8 ng/mL basic fibroblast growth factor (bFGF) as previously described (75). Colony initiating cell (CIC) assay was completed with a FACSAria II (BD); 1000 TRA-1-60+ and TRA-1-60– live cells were sorted into 6-well plates with irradiated MEF cells (iMEFs, seeded 24 h in advance), in serum replacement media, as previously described (68). For Oct4 staining, cells were incubated with Live/Dead Fixable Violet Dead Cell Stain 1:7000 (Invitrogen) for 30 min at 4°C, washed, fixed with Fix/Perm for 20 min at RT, washed with 1X Perm/Wash, incubated with Oct4-PE 1:100 (all from BD Biosciences) overnight, and washed thrice prior to whole well images acquired via automated high-content confocal fluorescence microscopy (Operetta, PerkinElmer). Germ layer differentiation was completed in MEF-CM for 4 days with daily media changes, with the following treatments, endoderm: Activin A 100ng/mL (R&D systems), CHIR99021 3uM (Sigma), hepatocyte growth factor 100ng/mL (Life Technology); Ectoderm: LDN193198 0.1uM (SelleckChem), SB431542 10uM (Sigma); and mesoderm: bone morphogenic protein-4 25ng/mL, FMS-like tyrosine kinase 3 ligand 300 ng/mL, and stem cell factor 100 ng/mL. For all experiments, tri12 and isogenic hPSCs were tested in parallel.

Cytogenetics

CLL patient sample cytogenetics were tested in hospitals, using fluorescence *in situ* hybridization for del(11)q, del(13)q, del(17)p, and acquisition of trisomy 12. Additional karyotyping and comparative genomic hybridization using Cytoscan HD Array (ThermoFisher Scientific) were performed by The Centre for Applied Genomics, at The Hospital for Sick Children, Toronto, Ontario, Canada. Cytoscan analysis was performed using Chromosome Analysis Suite (NetAffx 33.1, hg19) and default settings.

Differential gene expression and survival analysis

Total RNA was extracted from human PSCs using the RNeasy Micro Kit (Qiagen) following manufacturer's protocol. RNA was analyzed with the Affymetrix Human Gene 2.0 ST microarray (London Regional Genomics Centre, London, Ontario, Canada). Differential gene expression analysis was conducted using Partek Gene Suite (Partek v6.6), and included the original CA2 study (GEO accession: GSE10809, CA2 NoCre) (44) in addition to original data (GEO accession: GSE133697). In our previous study (40), human patients with clinical MBL and Rai 0 CLL donated PB samples, and the transcriptome of leukemic B cells was analyzed and is publicly available (GEO accession: GSE40570). Genes were considered differentially regulated with fold change >|2| and false discovery rate (FDR)-corrected $P<0.05$. Gene set enrichment analysis (GSEA, Broad Institute) was performed with default settings. Leading edge genes were individually mapped from 0 and 1 using a linear transformation which subtracted the minimum and divided by the maximum for N=22 patients, which was done to assist with interpretation and integration of the relative expression of each gene into a single 'leading edge genes' sum. These N=22 CLL and MBL patients (N=9 tri12, N=13 normal

cytogenetics, all *IGHV*-unmutated) were used for survival analysis as they received a minimum of 3 years of continuous disease monitoring. Patients were stratified based on LE gene expression into two equal-sized groups of N=11, survival differences were estimated with R packages *survival* and *survminer* (Kaplan–Meier method); built using *survfit* function, and plotted using *ggsurvplot* function. To assess the statistical significance of the survival stratification observed between patients with higher and lower LE gene expression, we computed a log-rank P value with *survdiff* function, using the R package *survival*. To assess the statistical significance of feature coefficients in multivariate Cox Proportional Hazards model, we computed each feature's hazard ratio (HR), 95% confidence intervals, and P value with *coxph* function, and plotted with *ggforest*, using the R package *survival*. A positive coefficient indicates a worse prognosis.

Tri12 Score derived using OCLR machine learning

To calculate a tri12 scoring index based on mRNA expression, we built a predictive model using one-class logistic regression (OCLR) as described by Malta et al. (47), and R script was adapted from deposited code (https://bioinformaticsfmrp.github.io/PanCanStem_Web/). Briefly, to ensure compatibility across datasets we selected only genes commonly shared across platforms, and the resulting training matrix consisted of 17,318 mRNA expression values common to all samples. Data was mean-centred and then OCLR was applied to only the tri12 hPSC samples. We validated our approach using leave-one-out cross-validation by withholding one tri12 hPSC sample in each iteration, and a separate signature was then trained on the other tri12 hPSC samples and used to score the withheld sample as well as the non-tri12 samples. Each iteration was measured using the area under the curve (AUC) metric, which can be interpreted as the probability that the model correctly ranks a positive sample above a negative sample. Every withheld tri12 hPSC sample was scored higher than all the non-tri12 samples, yielding an overall AUC of 1.0. The Tri12 Score was then applied to score new samples by computing Spearman correlations between the model's weight vector and the new sample's expression profile (which were accessed from GEO, NCBI: GSE40570) (40). The tri12 indices were then mapped from 0 and 1 using a linear transformation which subtracted the minimum and divided by the maximum. Mapping was done to assist with interpretation and integration of datasets.

B cell differentiation

TC-treated 10 cm dishes were coated with 0.1% gelatin (Sigma) for 1 h at 37°C prior to seeding stromal cells. MS5 cells were irradiated (iMS5) to prevent growth, and cryopreserved until use. OP9 cells were thawed and seeded at $4.0x10^5$ cells per 10 cm dish one day prior to hPSC harvest, in OP9 media: αMEM supplemented with 10% FBS (HyClone), 1 mM non-essential amino acids (Gibco), 1 mM L-glutamine (Gibco), and 100 μM monothioglycerol (Sigma). Human PSC were seeded as clumps onto OP9 cells; half of one 6-well of hPSC per 10 cm dish. Half media changes (10 mL) were performed on day 2 and 4 with OP9 media. One day prior to day 8 harvest, iMS5 cells were thawed and seeded at a density of 8.0x105 per 10 cm dish in iMS5 media; the same as OP9 media except FBS was previously heat inactivated (2 h at 56°C). Cells were harvested on day 8 by 20 min collagenase IV and 10 min 0.05% trypsin, keeping all non-adherent cells throughout. Day 8 single cells were stained with CD34 microbeads (Miltenyi) and enriched with LS columns following manufacturer's protocol. Purity of isolated CD34+ cells was >80%. CD34-enriched cells were seeded on to iMS5 cells

at a density of $3.0x10^5$ per 10 cm dish in iMS5 media, supplemented with 50 ng/mL stem cell factor (SCF, Sigma), 50 ng/mL FMS-like tyrosine kinase 3 ligand (FLT3L, Sigma), 20 ng/mL Interleukin-7 (IL-7, PeproTech), and 10 ng/mL IL-3 (R&D Systems). 10 mL half media changes occurred every 5 days for a total of 35 days, and SCF, FLT3L, and IL-7 were added each time. Day 42 cells were collected as described for day 8, and assessed by flow cytometry.

High-content drug testing

96-well plates of hPSCs were stained with Live/Dead Fixable Far Red Dead Cell Stain 1:7000 (Invitrogen, 30 min at 37°C), fixed using Fix/Perm Buffer (ThermoFisher Scientific, 20 min at RT), and nuclei were stained with Hoechst 33342 1:7000 (ThermoFisher Scientific) 10 min at RT. CLL cells were stained with 1 µg/mL calcein green in AIM-V media (both from ThermoFisher Scientific, 1 hour at RT). Whole 96-wells were evaluated for cytotoxicity by automated high-content confocal fluorescence microscopy (Operetta, PerkinElmer). Acquired images were analyzed using Acapella 2.0 (PerkinElmer) and FCS Express (De Novo Software v5) software. We thank Dr. Rima Al-war from the Ontario Institute of Cancer Research for supplying the OICR kinase library. Additional compounds were purchased as part of a chemical library from PerkinElmer (Guelph, ON).

Flow cytometry

Cells were stained using antibodies in the Table below. For intracellular staining, cells were incubated with Live/Dead Fixable Violet Dead Cell Stain 1:7000 (Invitrogen, 30 min at 37°C), stained with cell surface antibodies (if applicable, 30 min at 4°C), washed, fixed using Fix/Perm Buffer (20 min at RT), stained with antibodies overnight in Perm/Wash Buffer, and then washed thrice prior to acquisition. Live cells were stained for 30 min at 4°C in PBS with 3% FBS and 1 mM EDTA (Invitrogen), washed, and analyzed with an LSRII or Aria II flow cytometers (BD Biosciences). 7AAD 1:50 (Beckman Coulter) or Live/Dead Fixable Violet Dead Cell Stain was used to exclude nonviable cells. Fluorescence minus one (FMO; if applicable, secondary antibody but no primary antibody), flow controls were used to set gates in FlowJo (TreeStar, v10).

Antigen	Clone	Fluorochrome	Dilution	Catalogue	Supplier
Annexin	not applicable	FITC	1:100	556419	BD Biosciences
CD3	UCHT1	PE	1:100	IM1282U	Beckman Coulter
CD5	UCHT2	v450	1:100	561154	BD Biosciences
		APC-Cy7	1:1000	563516	BD Biosciences
CD19	HIB19	FITC	1:100	555412	BD Biosciences
CD20	2H7	PE	1:100	556633	BD Biosciences
CD23	M-L233	PE-Cy7	1:1000	561167	BD Biosciences
CD34	581	APC	1:100	555824	BD Biosciences
CD45	2D1	v450	1:100	642275	BD Biosciences
		APC	1:100	340943	BD Biosciences
	H130	PE	1:100	555483	BD Biosciences
CD79a	HM47	PerCP-C5.5	1:100	341643	BD Biosciences
CD79b	CD3-1	APC	1:100	550955	BD Biosciences
KIT	104D2	PE-Cy7	1:100	339195	BD Biosciences
EDNRB	polyclonal	purified	1:100	AF4496-SP	R&D Systems
FOLR1	548908	PE	1:100	FAB5646P	R&D Systems
IRAK4	L29-525	PE	1:100	560303	BD Biosciences

Antigen	Clone	Fluorochrome	Dilution	Catalogue	Supplier
Kappa	G20-193	APC-H7	1:100	561325	BD Biosciences
Lambda	1-155-2	PE	1:100	642919	BD Biosciences
OCT4	40/Oct-3	PE	1:100	560186	BD Biosciences
PAX5	1H9	PE	1:100	12-9918-80	ThermoFisher Scientific
TRA-1-60	TRA-1-60	Alexa Fluor 488	1:500	560173	BD Biosciences
		Alexa Fluor 647	1:1000	560122	BD Biosciences
TRA-1-85	TRA-1-85	FITC	1:100	560380	BD Biosciences

Xenotransplantation

Immunodeficient NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG) mice were bred in a barrier facility and all experimental protocols were approved by the Animal Research Ethics Board of McMaster University. Mouse sex and age were controlled within each experiment, and mice were randomly assigned to experimental groups which included male and female mice. No statistical method was used to predetermine sample size. For teratoma experiments, representative 6-wells of confluent PSCs were counted, and determined 1/2 6-wells contained $2x10^6$ cells. PSCs were then collected as clumps and pelleted individually in Eppendorf tubes. 500 µL of Matrigel (Corning) was added on ice. NSG mice were injected subcutaneously (SQ) with these cell mixtures: tri12 and isogenic PSCs in the right and left flank, respectively. Tumors were not observed to migrate and stayed stationary during the entire experiment. At endpoint, tumors were excised and tumour mass was weighed. EDNRB-i (Bosentan) and PF06650833 (IRAK4-i) were administered in 10% DMSO in corn oil, at 15 mg/kg daily for 21 consecutive days. For human CLL PDX modelling, mice were sublethally irradiated (single dose of 315 cGy, 137Cs) 24 h before transplant. Cells were transplanted by intravenous injection with $100x10^6$ PB MNCs as previously described (54), from patient CLL#160. EDNRB-i (Bosentan) and PF06650833 (IRAK4-i) were administered in 10% DMSO in corn oil, at 100 mg/kg daily for 4 consecutive days, starting on the 4th day post-transplant. Mice were harvested on day 8 post-transplant and killed by cervical dislocation. PB was immediately collected using trunk bleeding into K2 EDTA tubes, mixed immediately by inverting three times, and then immediately analyzed on the blood analyzer (scil Vet abc Plus+). Spleens were dissected and cells were recovered by mechanical dissociation in IMDM supplemented with 3% FBS, and 1 mM EDTA. Red blood cells were lysed using ammonium chloride. Phenotyping was analyzed using FMO gating (refer to *Flow cytometry* section). Data points were combined from all independent experiments and outliers were not excluded.

Histology

Tumors were excised from mice, weighed after draining of cysts (if any), and cut longitudinally to increase surface area for fixation in 10% formalin for 72 h. Tissues were then washed with PBS, then 50% ethanol for 2 h, and submitted to the MIRC Histology Core Facility (McMaster University), for standard Paraffin processing and embedding (Leica ASP300 Enclosed Tissue Processor), and hematoxylin and eosin (H&E) slide staining (Leica RM2255 Automated Microtome, Leica Autostainer XL, and CV5030 Coverslipper).

Statistical analysis

Data are represented as means ± standard deviation (SD). Prism (GraphPad v7) software was used for all statistical analyses, unless otherwise indicated. The criterion for statistical significance was typically $P<0.05$, or more stringent at $P<0.01$ for chemical genomic drug

testing (Fig 3). All GI50 values were calculated using log(inhibitor) vs. response – Variable slope (four parameters) statistics, constrained between the values 0 (venetoclax, positive control) and 100 (0.1% DMSO, negative control). If the GI50 curve was not converged, GI50 is represented as the highest dose tested (i.e. 10 μM is log GI50 = 4). Z-factor scores for CLL patient drug testing were 0.69 ± 0.08, and for hPSC drug testing were 0.56 ± 0.05. A Z-factor of 0.5-1.0 is an excellent assay (GraphPad, Article #1153). Power of Analysis using two different calculations (Stat UBC and ClinCalc), determined 4 additional patient samples per group (tri12 and normal) would substantiate our claims, at a power of 0.95 and alpha 0.05 error rate. This was determined using the GI50 and variance values determined using tri12 and normal CLL patient B cells (CLL#1 and CLL#2). In total, we screened N=12 tri12 CLL patient B cell samples, and N=18 $FISH^{NEG}$ CLL patient B cell samples (Table 1). Selectivity index was calculated using the mean GI50 of tri12 CLL patient samples divided by the mean GI50 of $FISH^{NEG}$ CLL patient samples, where GI50 is the concentration required to kill 50% of the cell population. Additional statistics are described in Figure legends.

Acknowledgements

We thank the patients and their families for their participation. We thank D. Porras for valuable comments. **Funding:** J.C.R. was supported by the Canadian Institutes of Health Research Doctoral Award, an Ontario Graduate Fellowship, and Michael G. DeGroote Doctoral Scholarship. D.G. was supported by the Jans Graduate Scholarship. L.A. and A.N. were supported by Associazione Italiana Ricerca sul Cancro (AIRC IG16722, IG10136). This work was supported by research grants to M.B. from the Leukemia & Lymphoma Society of Canada (353141 and 567994) and support to M.B. from the Canada Research Chair Program (Tier 1) in Human Stem Cell Biology, and Michael G. DeGroote Chair in Stem Cell Biology. The Quebec Leukemia Cell Bank (Banque de cellules leucémiques du Québec, Montreal, Quebec, Canada), receives funds from the Cancer Research Network of the Fonds de Recherche du Québec Santé (FRQS). The Manitoba Tumour Bank (Winnipeg, Manitoba, Canada) receives funds from the CancerCare Manitoba Foundation and the Canadian Institutes of Health Research, and is a member of the Canadian Tissue Repository Network. **Author contributions:** J.C.R., D.G., L.O., and A.L.B. performed experiments. L.A. provided clinical annotation. M.G., L.A., F.M., and A.N. provided patient annotation and clinical expertise. J.H., and B.L. provided patient samples and clinical expertise. C.J.V.C. provided clinical expertise. J.C.R. and M.B. designed experiments, interpreted data, and wrote the manuscript. M.B. directed the study. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Gene expression data from this study have been deposited in GEO (NCBI) under the accession number GSE133697. CA2 hESC were obtained by MTA from Dr. A. Nagy, Lunenfeld-Tanenbaum Research Institute, Toronto, Ontario. PB MNC from CLL patients were obtained through MTAs from the Quebec Leukemia Cell Bank and Manitoba Tumour Bank.

Figures and Tables

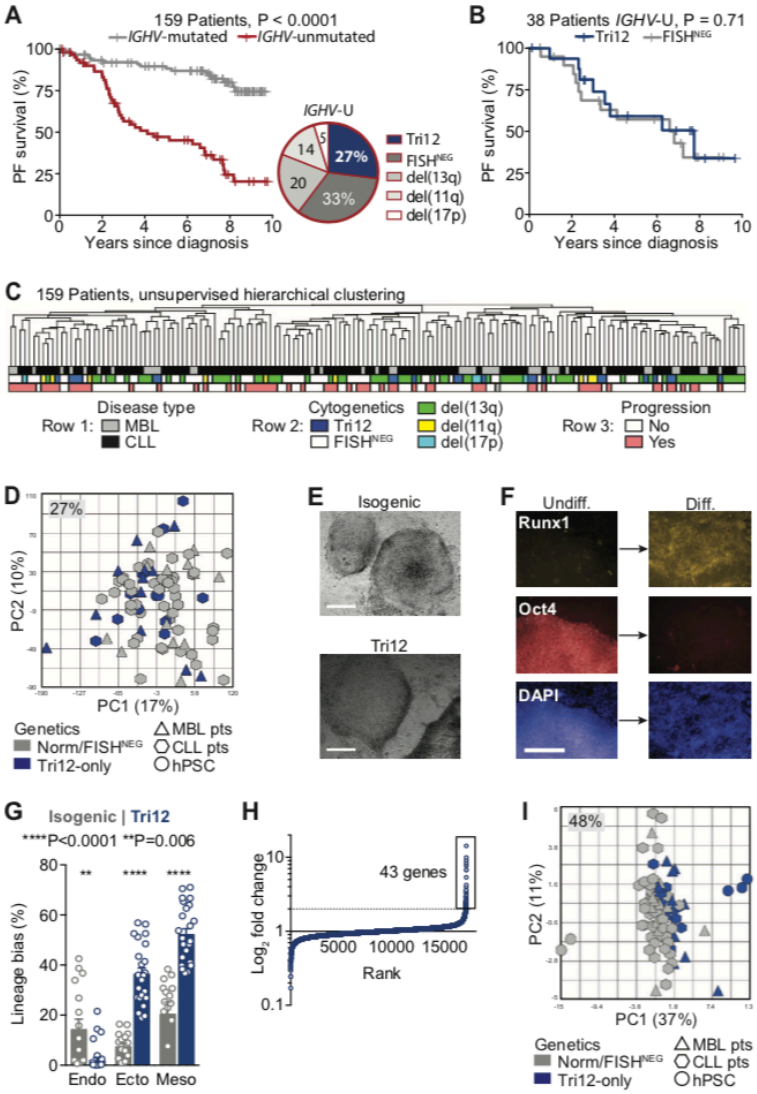


Figure 1. Pluripotent model removes confounding CLL interpatient heterogeneity.

(**A-B**) Kaplan-Meier estimator of progression free (PF) survival since diagnosis of 159 early stage CLL patients (**A**), and the most prevalent subtypes in cohort with significantly worse prognosis (**B**). Statistical assessment was performed using log rank test. (**C**) Unsupervised hierarchical clustering of global transcriptome of B cells purified from 159 CLL and MBL patients. Patient characteristics are described with colour coding. (**D**) Principal component (PC) analysis of tri12 and FISH^{NEG} patients from (**C**) using global transcriptome. (**E**) Morphology of whole colonies of undifferentiated hPSCs. Scale bar is 0.5 mm. (**F**) Representative staining of undifferentiated and differentiated hPSCs. Scale bar is 0.5 mm. (**G**) Germ layer marker expression relative to control. *N*=2 independent experiments *n*=16 biological replicates, each. Abbreviations: ectoderm, ecto; endoderm, endo; mesoderm, meso. Bars are mean ± SD. One-way ANOVA, **P<0.01, ****P<0.0001. (**H**) Global gene expression and selected top tri12-upregulated genes (ANOVA; fold change>2, FDR-P-value<0.05) from tri12 vs. isogenic hPSCs. (**I**) PC analysis of tri12 and FISH^{NEG} patients from (**C**) using isogenically controlled tri12 hPSCs (Table S2).

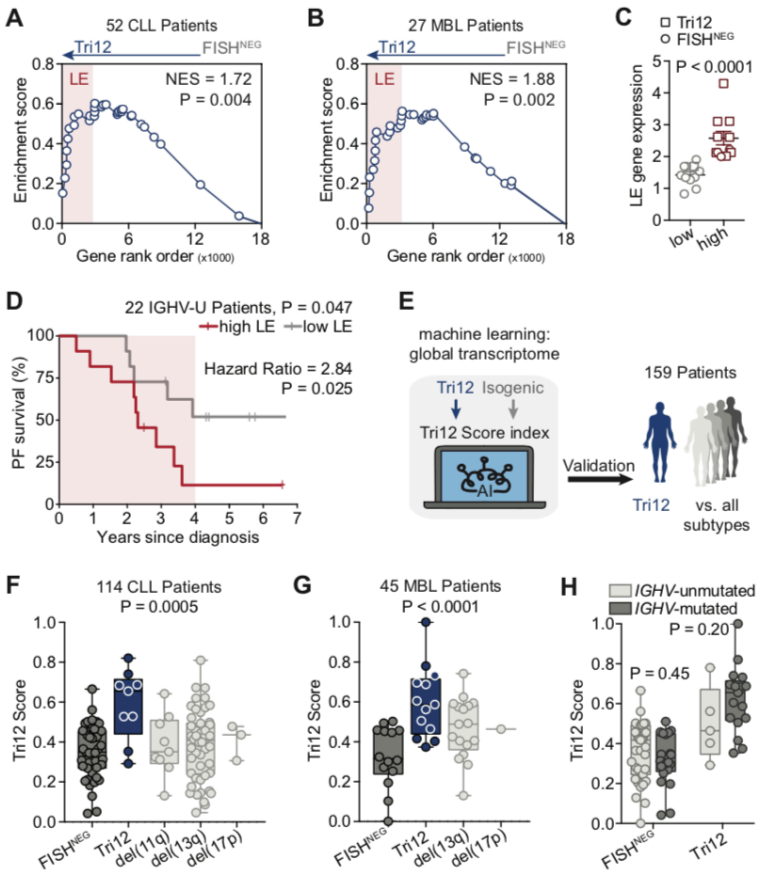


Figure 2. Tri12 gene networks underlie disease progression in patients.

(**A-B**) GSEA result (1,000 permutations) of 79 patients, segregated into early stage CLL (**B**) and clinical MBL (**C**). Genes on sex chromosomes removed. Leading edge (LE) in red. (**C-D**) LE gene expression in patients (**C**), and Kaplan-Meier estimator of PF survival (**D**) for 22 patients with >3 years of disease monitoring and disease progression within 4 years. Patients stratified into equal halves by LE gene expression. Statistical assessment was performed using two-tailed T-Test (**C**), log rank test, and Cox Proportional-Hazards model (**D**). (**E**) Schematic of machine learning using global transcriptomes. (**F-H**) Tri12 Score of CLL patients (**F**), MBL patients (**G**), and CLL and MBL patients subdivided by *IGHV* mutation status (**H**). Two-way ANOVA. Clinical features of patients in Table S2.

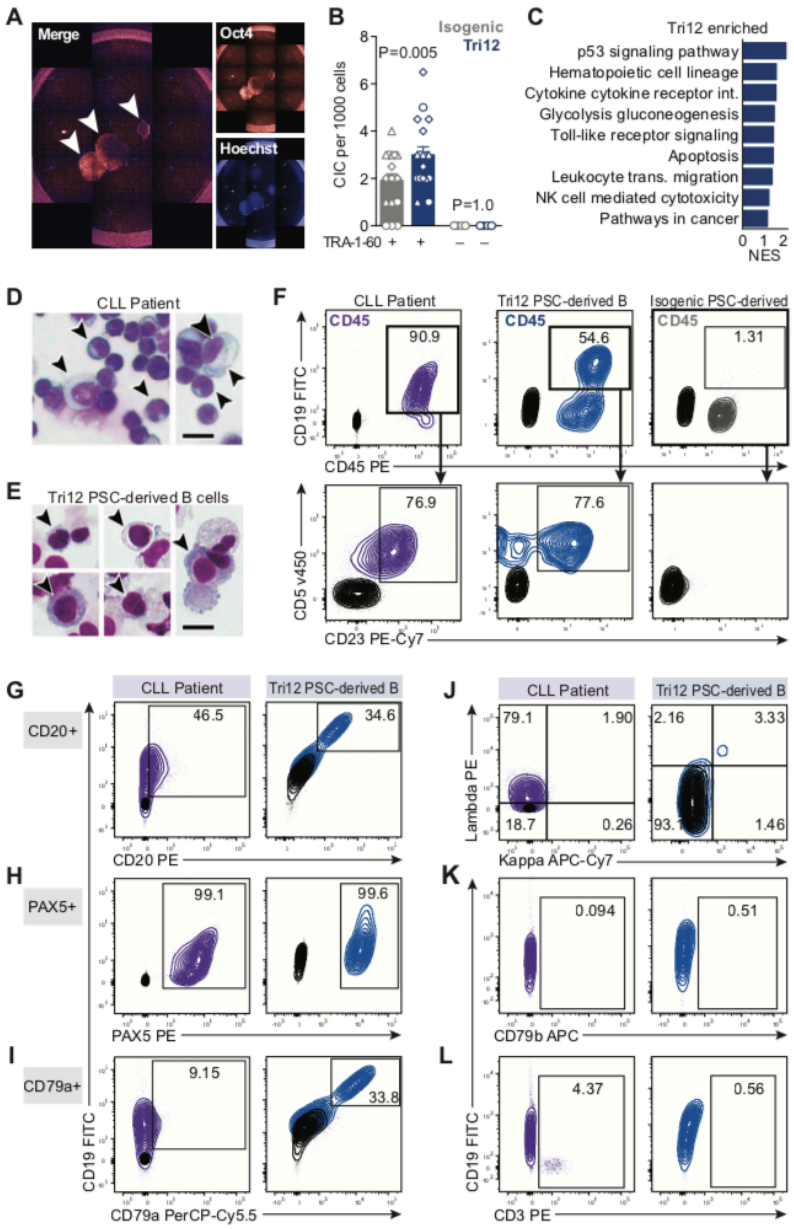


Figure 3. Cell autonomous role of tri12 in features of leukemogenesis.

(**A-B**) Oct4+ colony initiating cells (CIC; **A**), quantified per 1000 pluripotent (TRA-1-60+) or differentiated (TRA-1-60–) cells (**B**). N=3 independent experiments n=18 biological replicates. One-way ANOVA. (**C**) Result of GSEA (1,000 permutations) of KEGG gene sets significantly enriched in tri12 vs. isogenic hPSCs. NES, normalized enrichment score. FDR-P<0.2. (**D-E**) Morphology of CLL patient PB MNC (**D**) and B-lymphoid differentiation of tri12 hPSCs (**E**). (**F-K**) Flow cytometric analysis of malignant B-lymphopoiesis markers, in CLL patient PB (purple) and B-lymphoid differentiation of tri12 hPSCs (blue). Negative stain overlaid in black. N=7 independent experiments. (**L**) Flow cytometric validation of T-lymphoid absence.

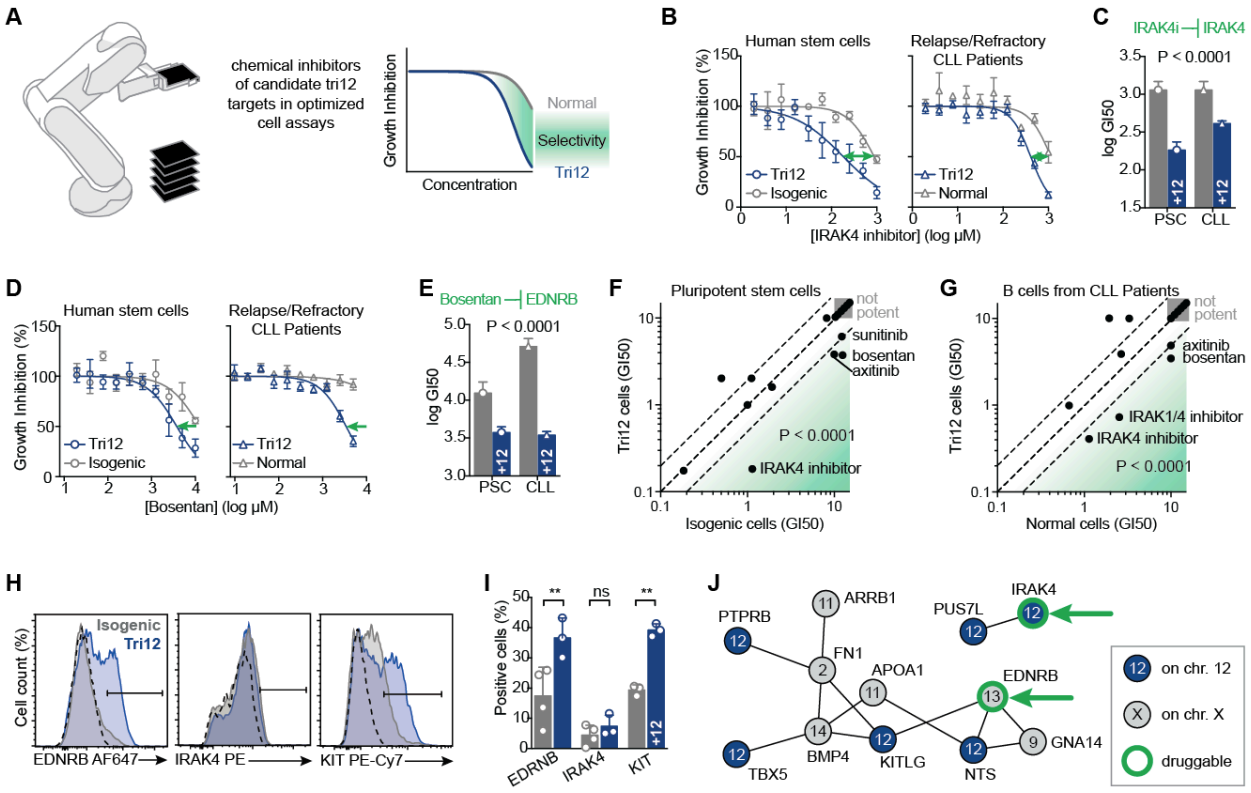


Figure 4. Identification of tri12-selective chemical inhibitors.

(**A**) Workflow of primary inhibitor screens using human PSCs and relapse/refractory CLL patient samples. (**B-C**) Growth inhibition results of IRAK4 inhibitor. Data points are mean±SD. N=2 independent experiments, n≥40 degrees of freedom. Green arrows indicate GI50. One-tailed T-Tests at 99% confidence. (**D-E**) Growth inhibition results of EDNRB inhibitor. Data points are mean±SD. N=2 independent experiments, n≥40 degrees of freedom. Green arrows indicate GI50. One-tailed T-Tests at 99% confidence. (**F**) Identification of tri12-sensitive chemical inhibitors in hPSC screen (green; tri12 GI50 < isogenic GI50 value). One-tailed T-Tests at 99% confidence. (**G**) Identification of tri12-sensitive chemical inhibitors in relapse/refractory CLL patient sample screen (green; tri12 GI50 < isogenic GI50 value). One-tailed T-Tests at 99% confidence. (**H-I**) Flow cytometry analysis (**H**), and quantification (**I**) of lead tri12 targets in hPSCs. N=3 independent experiments. One-way ANOVA. Source data and statistics for data in Table S6. (**J**) STRING plot showing tri12 up-regulated genes identified from isogenic-controlled tri12 transcriptome (Table S3). Disconnected nodes were removed. Chromosome (chr.) location is indicated in the centre of each node, genes located on chr. 12 are indicated in blue. Druggable genes with demonstrated tri12-selectivity are shown edged in green.

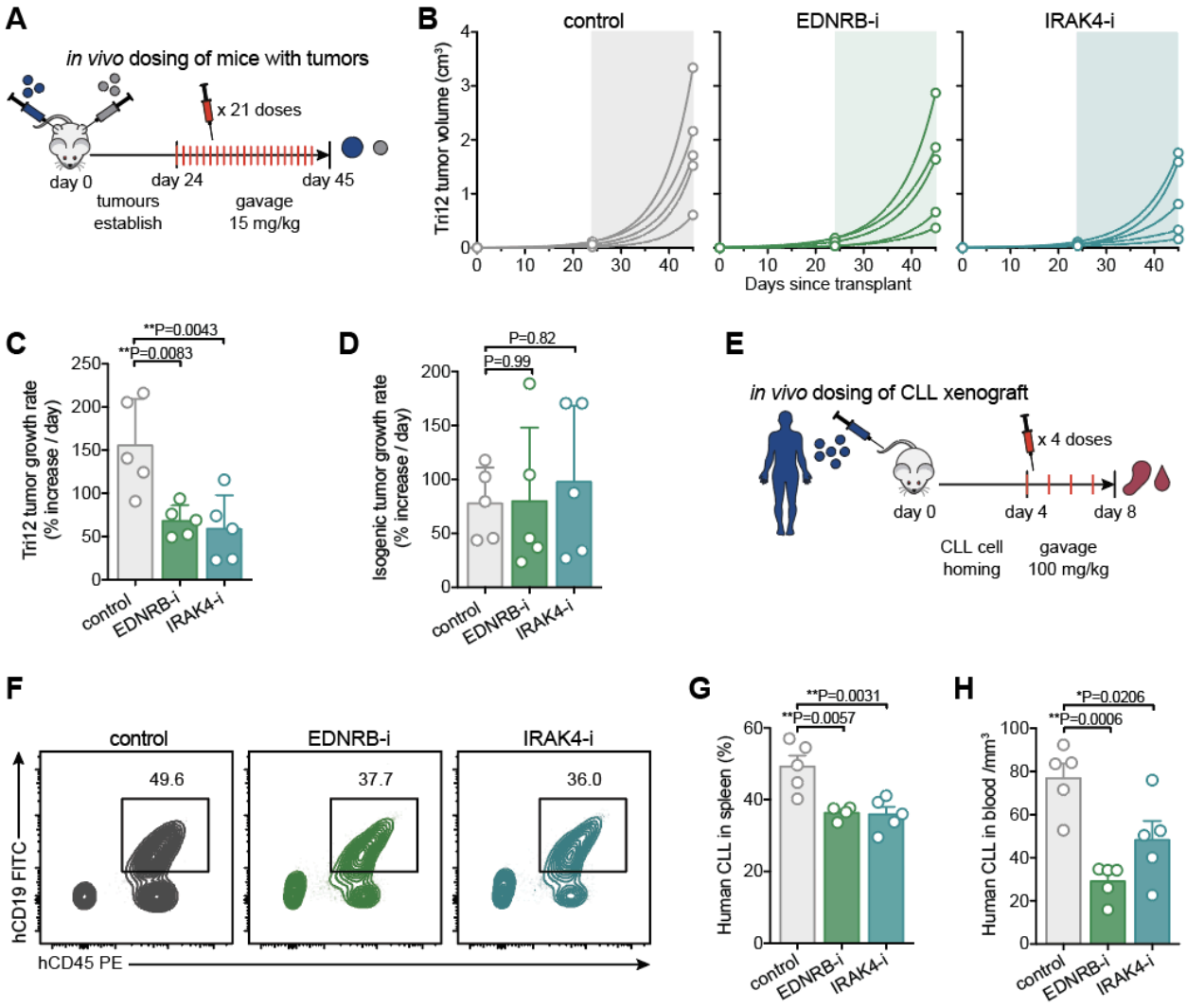
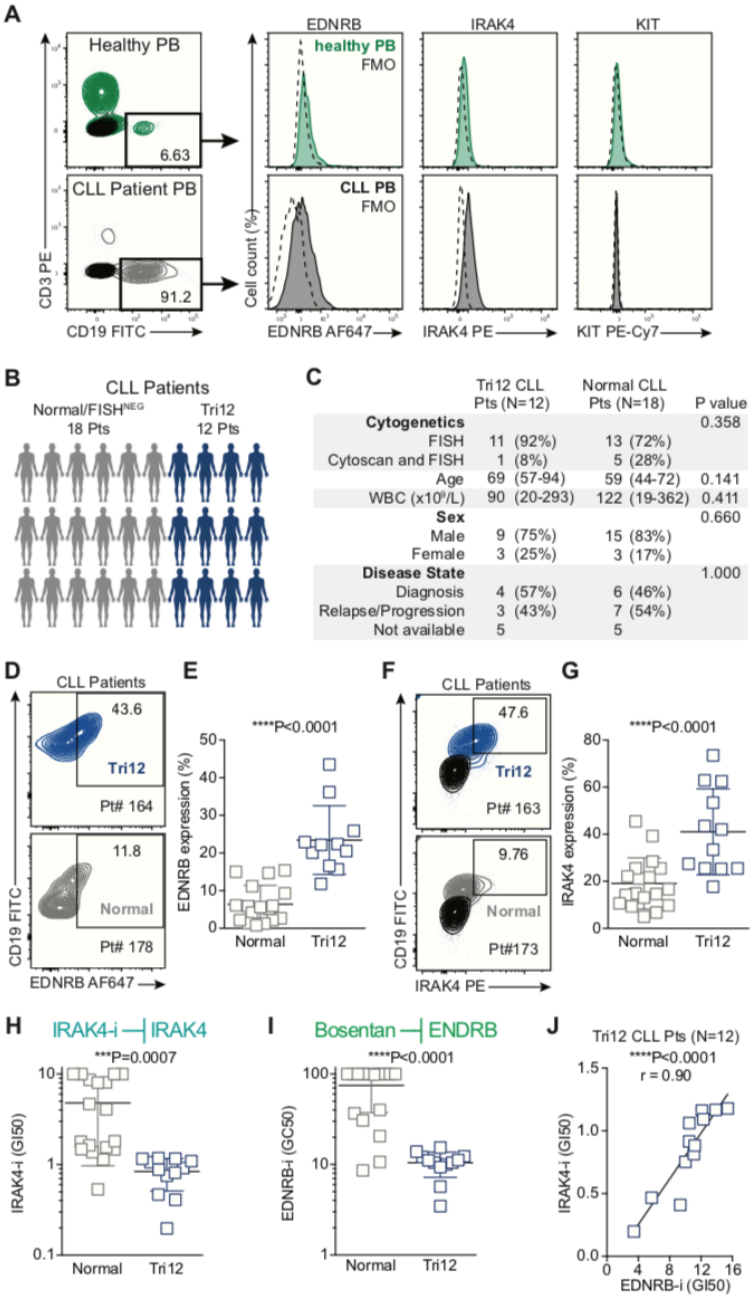


Figure 5. Chemically targeting tri12 tumor cells *in vivo*.

(**A**) Schematic depicting SQ transplantation of tri12 and isogenic PSCs into mice to establish tumors prior to daily oral gavage of two lead inhibitors for 3 weeks. (**B**) Tri12 PSC-derived tumor volume, depicted as exponential growth. Each curve is one mouse. Shaded area depicts when daily drug treatments occurred. (**C**) Tri12 PSC-derived tumor growth rate, comparing tumor volume prior to initiation of drug treatment and tumor volume at completion of drug treatment. Each dot is one mouse. One-way ANOVA. (**D**) Isogenic PSC-derived tumor growth rate, comparing tumor volume prior to initiation of drug treatment and tumor volume at completion of drug treatment. Each dot is one mouse. One-way ANOVA. (**E**) Schematic depicting IV transplantation of tri12 CLL into mice to establish engraftment prior to daily oral gavage of two lead inhibitors for 4 days. (**H**) Flow cytometry of live cells harvested from CLL PDX mouse spleens. (**I**) Human CLL chimerism (%CD45+CD19+) of live cells harvested from CLL PDX mouse spleens. One-way ANOVA. (**G**) The number of human CLL cells in freshly collected PB from CLL PDX mice (WBC count x frequency of CLL MNCs in WBCs). Each dot is one mouse. One-way ANOVA.



	Tri12 CLL Pts (N=12)	Normal CLL Pts (N=18)	P value
Cytogenetics			0.358
FISH	11 (92%)	13 (72%)	
Cytoscan and FISH	1 (8%)	5 (28%)	
Age	69 (57-94)	59 (44-72)	0.141
WBC (x10⁹/L)	90 (20-293)	122 (19-362)	0.411
Sex			0.660
Male	9 (75%)	15 (83%)	
Female	3 (25%)	3 (17%)	
Disease State			1.000
Diagnosis	4 (57%)	6 (46%)	
Relapse/Progression	3 (43%)	7 (54%)	
Not available	5	5	

Figure 6. Selective targeting of tri12 CLL patient cells.

(**A**) Protein expression of tri12 biomarkers in PB MNC from N=3 healthy adult volunteers and CLL. (**B**) Schematic of 30 CLL patient samples tested in (C-J). (**C**) Statistical assessment of 30 CLL patients was performed using Fisher's exact test. (**D-G**) Flow cytometric analysis of EDNRB (**D,E**) and IRAK4 (**F,G**) expression of 30 CLL samples. Two-tailed T-Tests. (**H**) IRAK4-i GI50 of 30 CLL samples. Two-tailed T-Test. (**I**) EDNRB-i (bosentan) GI50 of 30 CLL samples. Two-tailed T-Test. (**J**) Linear regression analysis of GI50 values, per CLL patient sample. CLL patients further described in Table 1.

Table 1. Thirty CLL patient samples used for drug screening.

Tri12 (grey shading) and normal (unshaded) CLL patient samples are unshaded. (*) Each PB sample was donated from a different CLL patient, except for two which are paired from the same CLL patient but separated by 7 years. (**) Source of CLL patient samples: Juravinski Hospital, Hamilton, Ontario (ON); Manitoba Tumour Bank, Winnipeg, Manitoba (MB); Quebec Leukemia Cell Bank, Montreal, Quebec (QC). (†) Experimental design of grouping samples for drug testing, which includes tri12 and normal CLL patient samples in all experiments. (†) Disease status (Dx, diagnosis; Prog., progression) and/or time from diagnosis until when the sample was collected, in years. (††) White blood cell and lymphocyte counts (x10^9 cells/L); normal WBC range is 1.3-4.4x10^9 cells/L (median 2.7x10^9 cells/L). (‡) Cytogenetics were assessed using 4 probe FISH or whole genome copy number using the Cytoscan HD microarray (Global). (‡‡) Total number of treatment regimens from Dx until donation. *Abbreviations*: mutated, M; patient, Pt; unmutated, U; not available, na; treatment, Tx.

CLL Pt#	Site **	Sex	Age	Status (Time) †	WBC ††	Rai Stage	CD38 (%)	Cytogenetics ‡		Tx No. ‡‡	*IGHV* Status
								Subtype	Assay		
160	ON	M	57	Prog. (9)	177	II	1	tri12	Global	5	na
161	ON	M	55	Prog. (17)	35	IV	3	normal	Global	7	na
162	MB	M	73	(4)	27	III	na	tri12	FISH	1	M
163	MB	F	71	(2)	36	II	61	tri12	FISH	1	M
164	MB	M	65	(4)	156	I	8	tri12	FISH	1	M
165	MB	M	94	(4)	51	0	12	tri12	FISH	1	U
166	MB	M	68	(2)	60	I	60	tri12	FISH	1	U
167	QC	M	63	Dx	20	na	3	tri12	FISH	na	na
168	QC	F	57	Dx	20	na	40	tri12	FISH	na	na
169*	QC	M	69	Dx	52	na	2	tri12	FISH	na	na
170	QC	M	73	Dx	293	na	2	tri12	FISH	na	na
171	QC	F	62	Prog.	58	na	3	tri12	FISH	na	na
172*	QC	M	76	Prog.	130	na	6	tri12	FISH	na	na
173	MB	F	66	(5)	232	I	0	normal	FISH	3	U
174	MB	M	71	(13)	19	I	42	normal	FISH	2	U
175	MB	F	63	(8)	19	I	14	normal	FISH	2	M
176	MB	M	72	(9)	80	0	0	normal	FISH	0	M
177	MB	M	61	(15)	111	0	0	normal	FISH	1	U
178	QC	M	46	Dx	29	na	1	normal	FISH	na	na
179	QC	M	44	Prog.	33	na	3	normal	FISH	na	na
180	QC	M	48	Dx	36	na	42	normal	FISH	na	na
181	QC	M	48	Prog.	108	na	0	normal	FISH	na	na
182	QC	M	56	Prog.	52	na	51	normal	FISH	na	na
183	QC	M	50	Dx	362	na	0	normal	FISH	na	na
184	QC	F	72	Prog.	53	na	2	normal	FISH	na	na
185	QC	M	52	Dx	335	na	51	normal	FISH	na	na
186	QC	M	72	Prog.	21	na	2	normal	Global	na	na
187	QC	M	67	Prog.	270	na	40	normal	Global	na	na
188	QC	M	53	Dx	222	na	1	normal	Global	na	na
189	QC	M	64	Dx	184	na	1	normal	Global	na	na

References

1. R. Beekman et al., The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nature Medicine, (2018).
2. A. Lin et al., Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. Science Translational Medicine 11, eaaw8412 (2019).
3. S. Parikh, Chronic lymphocytic leukemia treatment algorithm 2018. Blood Cancer Journal 8, (2018).
4. D. Rossi, V. Spina, G. Gaidano, Biology and treatment of Richter syndrome. Blood 131, 2761-2772 (2018).
5. M. Hallek, M. Furstenau, How to approach CLL in clinical practice. Hematol Oncol 37 Suppl 1, 38-42 (2019).
6. P. A. Thompson et al., Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival in IGHV-mutated chronic lymphocytic leukemia. Blood 127, 303-309 (2016).
7. T. Hilal, J. C. Gea-Banacloche, J. F. Leis, Chronic lymphocytic leukemia and infection risk in the era of targeted therapies: Linking mechanisms with infections. Blood Rev 32, 387-399 (2018).
8. K. Fischer et al., Venetoclax and Obinutuzumab in Patients with CLL and Coexisting Conditions. N Engl J Med 380, 2225-2236 (2019).
9. A. W. Roberts et al., Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. N Engl J Med 374, 311-322 (2016).
10. F. Awan et al., Acalabrutinib monotherapy in patients with chronic lymphocytic leukemia who are intolerant to ibrutinib. Blood Advances 3, 1553-1562 (2019).
11. J. C. Byrd et al., Acalabrutinib (ACP-196) in Relapsed Chronic Lymphocytic Leukemia. New England Journal of Medicine 374, 323-332 (2015).
12. D. A. Bond, J. A. Woyach, Targeting BTK in CLL: Beyond Ibrutinib. Current Hematologic Malignancy Reports (2019).
13. J. Brown, How I treat CLL patients with ibrutinib. Blood 131, 379-386 (2018).
14. B. L. Lampson et al., Idelalisib given front-line for treatment of chronic lymphocytic leukemia causes frequent immune-mediated hepatotoxicity. Blood 128, 195-203 (2016).
15. H. P. Byrd JC, O'Brien S, Barrientos JC, Reddy NM, Coutre S, Tam CS, Mulligan SP, Jaeger U, Barr PM, Furman RR, Kipps TJ, Thornton P, Moreno C, Montillo M, Pagel JM, Burger JA, Woyach JA, Dai S, Vezan R, James DF, Brown JR., Long-term follow-up of the RESONATE phase 3 trial of ibrutinib vs ofatumumab. Blood 133, 2031-2042 (2019).
16. A. Quinquenel et al., Prevalence of BTK and PLCG2 mutations in a real-life CLL cohort still on ibrutinib after three years: a FILO group study. Blood, blood.2019000854 (2019).
17. B. Druker et al., Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr–Abl positive cells. Nature Medicine, 561-566 (1996).
18. Z.-Y. Wang, Z. Chen, Acute promyelocytic leukemia: from highly fatal to highly curable. Blood 111, 2505-2515 (2008).
19. D. L. Van Dyke et al., The Dohner fluorescence in situ hybridization prognostic classification of chronic lymphocytic leukaemia (CLL): the CLL Research Consortium experience. Br J Haematol 173, 105-113 (2016).
20. M. Hallek et al., iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. Blood 131, 2745-2760 (2018).
21. E. Vendramini et al., KRAS, NRAS, and BRAF mutations are highly enriched in trisomy 12 chronic lymphocytic leukemia and are associated with shorter treatment-free survival. Leukemia (2019).
22. P. Strati, T. D. Shanafelt, Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. Blood 126, 454-462 (2015).

23. T. Hassold et al., A cytogenetic study of 1000 spontaneous abortions. Ann Hum Genet 44, 151-178 (1980).
24. J. Hu et al., Four children with postnatally diagnosed mosaic trisomy 12: Clinical features, literature review, and current diagnostic capabilities of genetic testing. Am J Med Genet A, (2020).
25. W. G. Finn, A. Sreekumar, A. Menon, C. Utiger, A. Chinnaiyan, Trisomy 12-associated, t(11;14)-negative mature B-cell leukemia with gene expression profile resembling mantle cell lymphoma. Leuk Lymphoma 47, 121-127 (2006).
26. F. Mitelman, G. Levan, Clustering of aberrations to specific chromosomes in human neoplasms. IV. A survey of 1,871 cases. Hereditas 95, 79-139 (1981).
27. S. Fiorcari et al., Overexpression of CD49d in trisomy 12 chronic lymphocytic leukemia patients is mediated by IRF4 through induction of IKAROS. Leukemia (2019).
28. L. V. Abruzzo et al., Trisomy 12 chronic lymphocytic leukemia expresses a unique set of activated and targetable pathways. Haematologica (2018).
29. B. Hu et al., Association of gene mutations with time-to-first treatment in 384 treatment-naive chronic lymphocytic leukaemia patients. Br J Haematol (2019).
30. A. J. Souers et al., ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nat Med 19, 202-208 (2013).
31. L. Honigberg et al., The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. Proceedings of the National Academy of Sciences of the United States of America 107, 13075-13080 (2010).
32. S. E. Herman et al., Phosphatidylinositol 3-kinase-delta inhibitor CAL-101 shows promising preclinical activity in chronic lymphocytic leukemia by antagonizing intrinsic and extrinsic cellular survival signals. Blood 116, 2078-2088 (2010).
33. M. S. Davids et al., Comprehensive Safety Analysis of Venetoclax Monotherapy for Patients with Relapsed/Refractory Chronic Lymphocytic Leukemia. Clinical Cancer Research 24, 4371-4379 (2018).
34. S. E. Coutre et al., Management of adverse events associated with idelalisib treatment: expert panel opinion. Leuk Lymphoma 56, 2779-2786 (2015).
35. A. Boyd et al., Identification of chemotherapy-induced Leukemic Regenerating Cells reveals a transient vulnerability of human AML recurrence. Cancer Cell 34, 483-498 (2018).
36. M. T. Bertilaccio et al., Xenograft models of chronic lymphocytic leukemia: problems, pitfalls and future directions. Leukemia 27, 534-540 (2013).
37. S. Oppermann et al., High-content screening identifies kinase inhibitors that overcome venetoclax resistance in activated CLL cells. Blood 128, 934-947 (2016).
38. S. Molica et al., Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. Am J Hematol 92 (2017).
39. D. D. Von Hoff et al., Pilot study using molecular profiling of patients' tumors to find potential targets and select treatments for their refractory cancers. J Clin Oncol 28, 4877-4883 (2010).
40. F. Morabito et al., Clinical monoclonal B lymphocytosis versus Rai 0 chronic lymphocytic leukemia: A comparison of cellular, cytogenetic, molecular, and clinical features. Clinical Cancer Research 19, 5890-5900 (2013).
41. D. Herszfeld et al., CD30 is a survival factor and a biomarker for transformed human pluripotent stem cells. Nature Biotechnology 24, 351-357 (2006).
42. T. E. Werbowetski-Ogilvie et al., Characterization of human embryonic stem cells with features of neoplastic progression. Nature Biotechnology 27, 91-97 (2009).
43. The International Stem Cell Initiative et al., Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nature biotechnology 29, 1132-1144 (2011).

44. The International Stem Cell Initiative et al., Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nature Biotechnology 25, 803-816 (2007).
45. M. Nakanishi et al., Human Pluripotency Is Initiated and Preserved by a Unique Subset of Founder Cells. Cell 177, 910-924 (2019).
46. H. Döhner et al., Genomic aberrations and survival in chronic lymphocytic leukemia. The New England Journal of Medicine 343, 1910-1916 (2000).
47. T. M. Malta et al., Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell 173, 338-354 (2018).
48. A. Sokolov, D. E. Carlin, E. O. Paull, R. Baertsch, J. M. Stuart, Pathway-Based Genomics Prediction using Generalized Elastic Net. PLoS Comput Biol 12, e1004790 (2016).
49. M. Gruber et al., Growth dynamics in naturally progressing chronic lymphocytic leukaemia. Nature 570, 474-479 (2019).
50. A. French, C. T. Yang, S. Taylor, S. M. Watt, L. Carpenter, Human induced pluripotent stem cell-derived B lymphocytes express sIgM and can be generated via a hemogenic endothelium intermediate. Stem cells and development 24, 1082-1095 (2015).
51. S. Martinelli et al., The expression of endothelin-1 in chronic lymphocytic leukemia is controlled by epigenetic mechanisms and extracellular stimuli. Leuk Res 54, 17-24 (2017).
52. L. Rosano, F. Spinella, A. Bagnato, Endothelin 1 in cancer: biological implications and therapeutic opportunities. Nat Rev Cancer 13, 637-651 (2013).
53. G. W. Rhyasen, D. T. Starczynowski, IRAK signalling in cancer. Br J Cancer 112 (2015).
54. A. Ljungars et al., A platform for phenotypic discovery of therapeutic antibodies and targets applied on Chronic Lymphocytic Leukemia. npj Precision Oncology 2, 18 (2018).
55. J. Durig et al., A novel nonobese diabetic/severe combined immunodeficient xenograft model for chronic lymphocytic leukemia reflects important clinical characteristics of the disease. Cancer Res 67, 8653-8661 (2007).
56. Y. Kikushige et al., Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. Cancer Cell 20, 246-259 (2011).
57. H. Hentze et al., in Stem cell research. (Elsevier B.V., 2009), vol. 2, pp. 198-210.
58. J. Chilloux et al., Microbiome Inhibition of IRAK-4 by Trimethylamine Mediates Metabolic and Immune Benefits in High-Fat-Diet-induced Insulin Resistance. bioRxiv, (2018).
59. K. G. Serafim et al., Bosentan, a mixed endothelin receptor antagonist, inhibits superoxide anion-induced pain and inflammation in mice. Naunyn Schmiedebergs Arch Pharmacol 388, 1211-1221 (2015).
60. S. Bien et al., The endothelin receptor blocker bosentan inhibits doxorubicin-induced cardiomyopathy. Cancer Research 67, 10428-10435 (2007).
61. A. K. Imhof et al., Potent anti-inflammatory and antinociceptive activity of the endothelin receptor antagonist bosentan in monoarthritic mice. Arthritis Res Ther 13, R97 (2011).
62. P. Hillmen et al., Ibrutinib Plus Venetoclax in Relapsed/Refractory Chronic Lymphocytic Leukemia: The CLARITY Study. Journal of Clinical Oncology, JCO.19.00894 (2019).
63. J. A. Burger et al., Long-term efficacy and safety of first-line ibrutinib treatment for patients with CLL/SLL: 5 years of follow-up from the phase 3 RESONATE-2 study. Leukemia (2019).
64. P. Bulian et al., Mutational Status of IGHV is the most reliable prognostic Marker in Trisomy 12 Chronic Lymphocytic Leukemia. Haematologica 102, e443-e446. (2017).
65. S. Decker et al., Trisomy 12 and elevated GLI1 and PTCH1 transcript levels are biomarkers for Hedgehog-inhibitor responsiveness in CLL. Blood 119, 997-1007 (2012).
66. A. Zucchetto et al., CD49d is overexpressed by trisomy 12 chronic lymphocytic leukemia cells: evidence for a methylation-dependent regulation mechanism. Blood 122, 3317-3321 (2013).
67. F. Autore et al., Elevated Lactate Dehydrogenase Has Prognostic Relevance in Treatment-Naive Patients Affected by Chronic Lymphocytic Leukemia with Trisomy 12. Cancers (2019).

68. J.-H. Lee et al., Brief Report: Human Acute Myeloid Leukemia Reprograming to Pluripotency Is a Rare Event and Selects for Patient Hematopoietic Cells Devoid of Leukemic Mutations. Stem Cells (2017).
69. K. Melgar et al., Overcoming adaptive therapy resistance in AML by targeting immune response pathways. Science Translational Medicine 11, eaaw8828 (2019).
70. S. Lee et al., Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. Nature 561, 127-131 (2018).
71. G. Cutrona et al., Microenvironmental regulation of the IL-23R/IL-23 axis overrides chronic lymphocytic leukemia indolence. Science Translational Medicine 10 (2018).
72. J. C. Byrd, Targeting CD20 takes the backseat in CLL. Blood 133, 1003-1004 (2019).
73. M. Skarzynski et al., Interactions between Ibrutinib and Anti-CD20 Antibodies: Competing Effects on the Outcome of Combination Therapy. Clinical Cancer Research 22, 86-95 (2016).
74. J. Burger et al., Randomized trial of ibrutinib vs ibrutinib plus rituximab in patients with chronic lymphocytic leukemia. Blood 133, 1011-1019 (2019).
75. J. C. Reid et al., CXCL12/CXCR4 Signaling Enhances Human PSC-Derived Hematopoietic Progenitor Function and Overcomes Early In Vivo Transplantation Failure. Stem Cell Reports 10, 1625-1641 (2018).
76. M. A. Vodyanik, J. A. Bork, J. A. Thomson, I. Slukvin, Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. Blood 105, 617-626 (2005).
77. Z. Li et al., Inhibition of IRAK1/4 sensitizes T cell acute lymphoblastic leukemia to chemotherapies. J Clin Invest 125, 1081-1097 (2015).
78. G. W. Rhyasen et al., Targeting IRAK1 as a therapeutic approach for myelodysplastic syndrome. Cancer Cell 24, 90-104 (2013).

Supplemental Figures and Tables

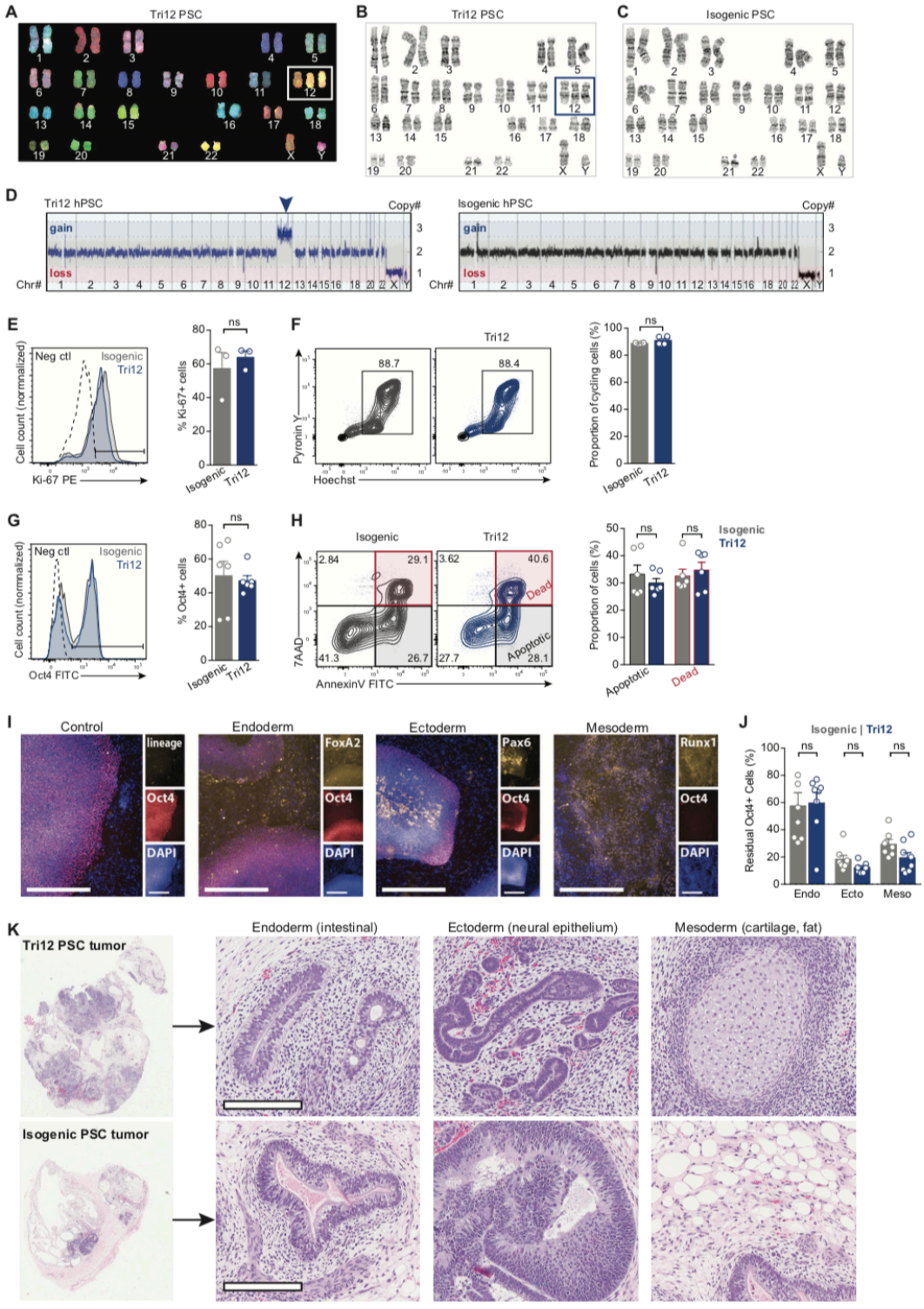


Figure legend on next page

Figure S1. Tri12 and isogenic hPSCs share pluripotent features.

(**A-B**) Spectral karyotype (**A**) and G-band karyotype (**B**) shows hPSCs with tri12 (50/50 metaphases). (**C**) Normal karyotype in isogenic hPSCs (50/50 metaphases). (**D**) Detection of DNA micro-insertions and deletions. Chromosome numbers along bottom, and DNA copy number along right edge. Arrow indicates chromosome 12. (**E**) Flow cytometry and quantification of hPSCs stained for Ki-67+ proliferating cells. *N*=3 independent experiments *n*=3 biological replicates, each. Two-tailed T-Test. FMO in black. (**F**) Flow cytometry and quantification of cell cycle status indicating hPSCs in S-G2-M phase. *N*=2 independent experiments *n*=4 biological replicates, each. Two-tailed T-Test. Negative control in black. (**G**) Flow cytometry and quantification of OCT4+ pluripotent cells. *N*=6 independent experiments *n*=6 biological replicates, each. Two-tailed T-Test. (**H**) Flow cytometry and quantification of apoptotic (AnnexinV+7AAD–) and dead (AnnexinV+7AAD+) hPSCs. *N*=2 independent experiments *n*=6 biological replicates, each. Two-tailed T-Test. (**I**) Representative staining on differentiation day 4. Scale bar is 0.5 mm. (**J**) Residual Oct4 expression relative to control, on day 4. *N*=1 independent experiments *n*=8 biological replicates, each. Bars are mean ± SD. One-way ANOVA. (**K**) Hematoxylin and eosin-stained sections of teratomas formed 7 weeks after subcutaneously injecting immunocompromised mice with tri12 and isogenic PSC. Individual teratomas contain all three embryonic germ layers: endoderm (intestinal epithelium), mesoderm (cartilage, skeletal muscle, adipose tissue) and ectoderm (neural epithelium with rosettes). Scale bars, 200 mm.

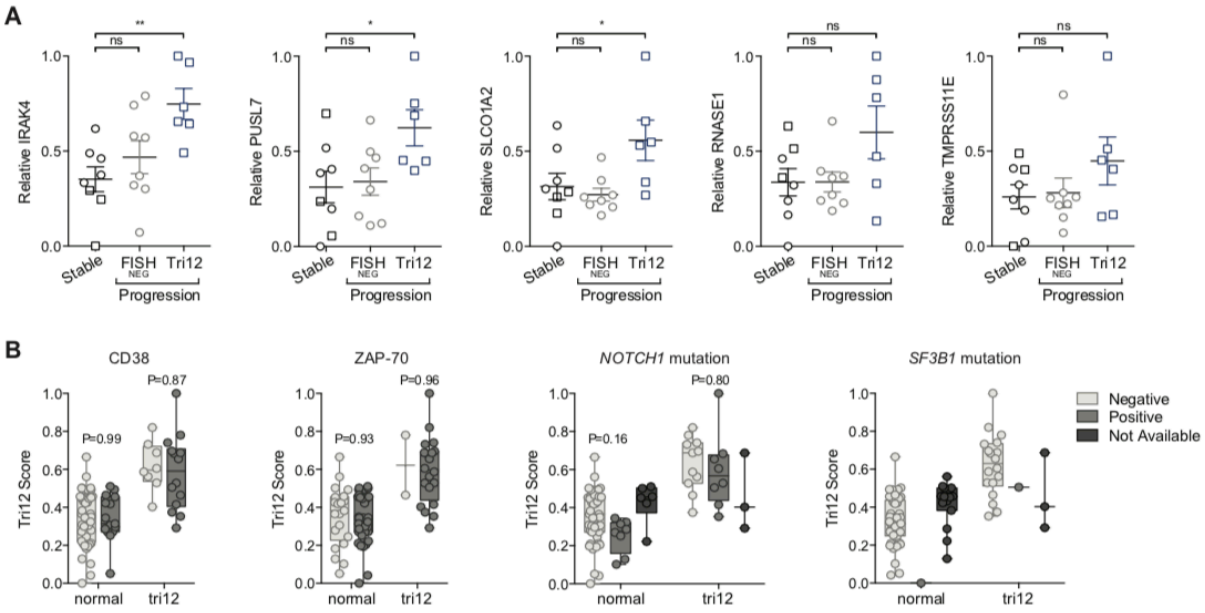


Figure S2. Isogenic-controlled tri12 genes transcend patient heterogeneity.

(**A**) Relative gene expression of individual genes common to leading edge (LE) of tri12 CLL and tri12 MBL. 'Stable' disease is progression-free (PF) status for >3 years, 'Progression' occurred within 4 years: tri12, squares; FISH^{NEG}, circles. (**B**) Tri12 Score shown for MBL and CLL patients (also in Table S1), segregated by additional genetic and phenotypic traits. Two-way ANOVAs conducted for Positive vs. Negative, except *SF3B1* mutation due to low sample sizes; P-values in Fig.

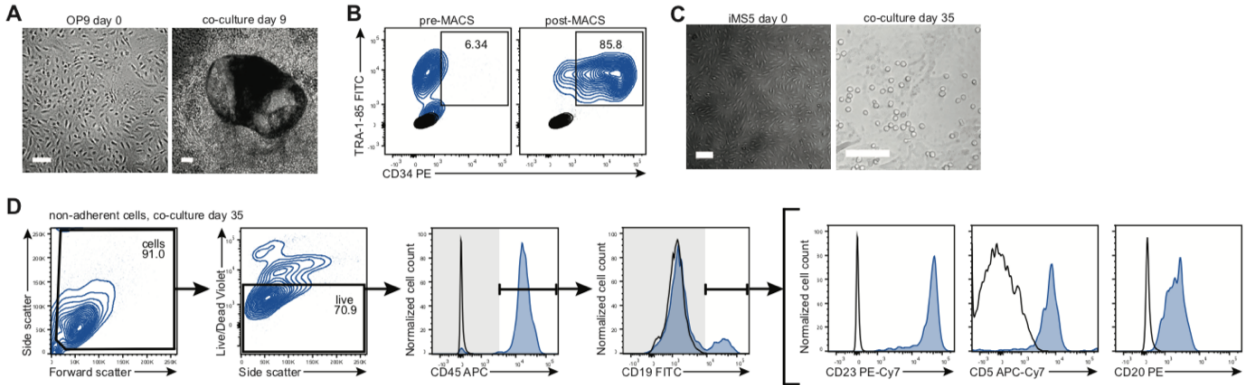


Figure S3. Extended assessment of B-lymphoid development from tri12 hPSCs.

(A) Representative images of OP9 monolayer and co-culture with tri12 hPSCs. Scale bar is 200 µm. (B) Representative enrichment of CD34+ cells from OP9 day 9 co-culture with tri12 hPSCs. (C) Representative images of irradiated MS5 (iMS5) monolayer and co-culture with tri12 hPSCs. Scale bar is 200 µm. Non-adherent round cells are observed upon prolonged co-culture. (D) Representative gating of non-adherent cells (collected separately) from iMS5 day 35 co-culture, with tri12 hPSC-derived B lymphoid derived cells. FMO in black.

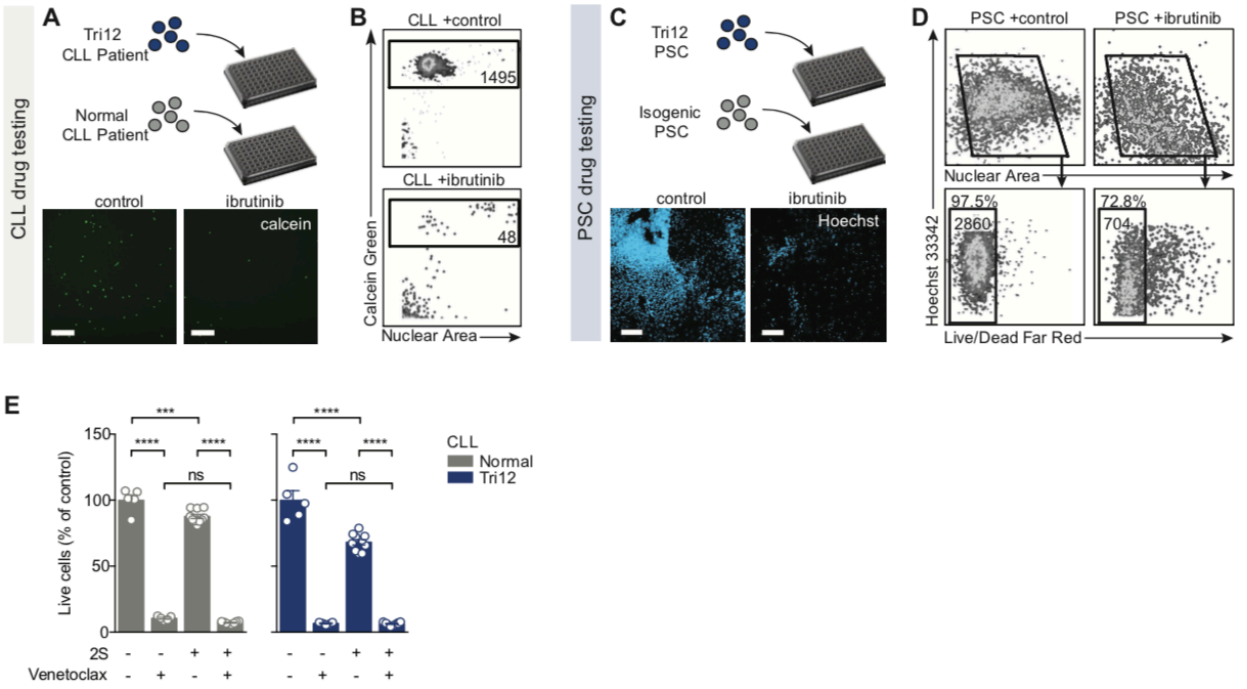


Figure S4. Optimized high-content drug testing of tri12 targets.

(**A**) Schematic of CLL patient drug testing in 96-wells, and representative live cells positively stained with calcein green. (**B**) Representative gating of CLL cells treated with control (0.05% DMSO) or ibrutinib. (**C**) Schematic of hPSC drug testing in 96-wells, and representative Hoechst staining of fixed hPSCs. (**D**) Representative gating of hPSCs treated with control (0.05% DMSO) or ibrutinib. (**E**) 2S conditions were evaluated for proliferation and venetoclax resistance with two CLL patient samples (CLL#160 and CLL#161). Each dot is a biological replicate. Two-way ANOVA, ***P<0.001, ****P<0.0001.

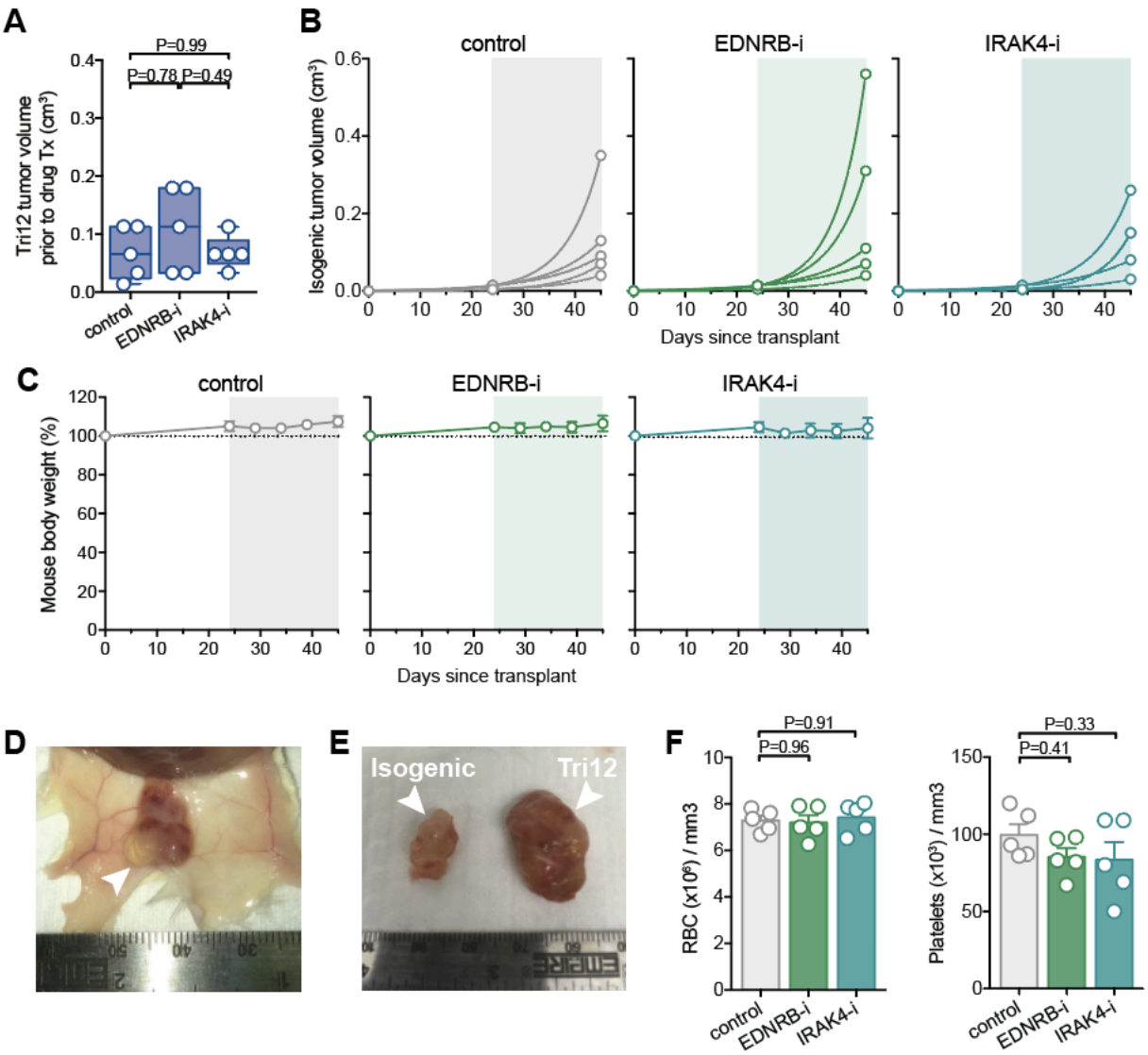


Figure S5. Extended assessment of PSC-derived tumours in mice.

(**A**) Individual tri12 PSC-derived tumor volumes are shown at 24 days post-transplant, prior to initiation of drug treatment. Each dot is one mouse. One-way ANOVA. (**B**) Isogenic PSC-derived tumor volume depicted as exponential growth. Each curve is one mouse. Shaded area depicts when daily drug treatments occurred. (**C**) Body weight of experimental mice normalized to their original start weight; shading indicates when oral gavage was administered. (**D**) Representative image of SQ tumor in a mouse at 45 days post-transplant. (**E**) Representative images of isogenic and tri12 PSC-derived tumors, excised from a mouse at 45 days post-transplant. (**F**) Freshly collected PB from CLL PDX mice was quantified with a blood analyzer. Each dot is one mouse. One-way ANOVA.

Table S1. Preclinical therapeutic potential pipeline for CLL.

Drug	Demonstration of therapeutic potential for CLL patients	Reference
venetoclax (ABT-199) first-in-class BCL2 inhibitor	cell lines in mice: • Granta-519: mantle cell lymphoma • DoHH2: non-Hodgkin's lymphoma with complex karyotype • Toledo: diffuse large cell lymphoma with translocation t(14;18) • RS4;11: acute lymphoblastic leukemia in patients: • 3 refractory CLL patients, 1 dose	Souers et al. *Nature Medicine* 2013.
ibrutinib (PCI-32765) first-in-class BTK inhibitor	cell line *in vitro*: • DoHH2: non-Hodgkin's B-cell lymphoma with complex karyotype *in vitro*: • human B cells, undisclosed patient characteristics and sample size mouse models: • collagen-induced arthritis • MRL-Fas(lpr) mice; arthritis, lupus, aberrant T cell proliferation dog model: • spontaneous non-Hodgkin lymphoma	Honigberg et al. *PNAS* 2010.
idelalisib (CAL-101) first-in-class PI3Kδ inhibitor	*in vitro*: • protein expression of CLL patient samples • drug treatment of CLL patient samples • drug treatment of CLL patient samples in the presence of a HPV-transformed human HS-5 fibroblast cell line	Herman et al. *Blood* 2010.

Table S2. MBL and CLL patient characteristics and Tri12 Score.

NA = not available; 0 = absent and 1 = present, M = mutated and U = unmutated.

GEO ID	Patient ID	Disease	Cytogenetics	IGHV	CD38	ZAP70	NOTCH1	SF3B1	Tri12 Score
GSM996784	AA0284	CLL	del13	M	0	0	0	0	0.67085
GSM996785	AB0350	CLL	del13	M	0	0	0	0	0.04654
GSM996786	AG0136	CLL	tri12	M	1	1	NA	NA	0.2917
GSM996787	AG0268	MBL	tri12	M	1	0	0	0	0.78021
GSM996788	AG0318	CLL	tri12	U	0	1	0	0	0.82052
GSM996789	AG0403	CLL	del13	M	0	1	0	0	0.58139
GSM996790	AM0034	CLL	normal	U	1	1	0	0	0.29728
GSM996791	AM0067	CLL	normal	U	0	1	0	0	0.30429
GSM996792	AM0220	CLL	tri12	U	1	1	1	0	0.35324
GSM996793	AM0270	CLL	del13	M	0	1	0	0	0.24336
GSM996794	AR0242	CLL	tri12	U	1	1	1	0	0.68410
GSM996795	AS0012	CLL	del13	M	0	0	0	0	0.67274
GSM996796	BB0245	CLL	normal	M	0	0	0	0	0.45287
GSM996797	BE0192	CLL	tri12	U	0	1	1	0	0.52897
GSM996798	BF0316	CLL	del13	M	0	1	1	0	0.43056
GSM996799	BL0162	MBL	tri12	M	0	1	0	0	0.56182
GSM996800	BP0326	CLL	tri12	U	1	1	1	0	0.65555
GSM996801	BR0218	MBL	del13	M	0	0	0	0	0.40504
GSM996802	BS0214	MBL	normal	M	0	1	0	0	0.19367
GSM996803	BV0176	MBL	del13	M	0	0	0	0	0.13031
GSM996804	BV0229	CLL	normal	M	0	0	0	0	0.18266
GSM996805	CA0058	CLL	normal	U	0	1	0	0	0.34962
GSM996806	CA0101	CLL	tri12	U	1	1	NA	NA	0.68720

GEO ID	Patient ID	Disease	Cytogenetics	IGHV	CD38	ZAP70	NOTCH1	SF3B1	Tri12 Score
GSM996807	CA0160	CLL	del13	M	0	0	0	NA	0.36482
GSM996808	CA0345	CLL	del13	M	0	0	0	0	0.54846
GSM996809	CC0244	CLL	del13	M	0	1	0	0	0.49828
GSM996810	CC0311	MBL	del13	M	0	1	0	0	0.57550
GSM996811	CD0310	CLL	del13	U	1	1	0	0	0.08609
GSM996812	CE0078	MBL	del13	M	0	1	NA	NA	0.48873
GSM996813	CE0399	CLL	del13	U	0	1	1	0	0.80982
GSM996814	CG0166	CLL	normal	M	0	0	0	0	0.24531
GSM996815	CG0204	CLL	del13	M	0	0	0	NA	0.33610
GSM996817	CM0057	CLL	del11	U	1	1	0	0	0.35150
GSM996818	CM0246	CLL	del13	M	0	1	0	0	0.21866
GSM996819	CM0289	CLL	del11	U	0	1	0	0	0.13099
GSM996820	CP0104	CLL	del13	U	0	1	0	0	0.41213
GSM996821	CR0203	CLL	del17	U	1	1	1	0	0.43603
GSM996822	CR0297	CLL	tri12	U	1	1	0	0	0.52762
GSM996823	CS0030	MBL	tri12	U	0	1	1	0	0.60504
GSM996824	CS0290	CLL	del17	M	0	0	0	NA	0.30705
GSM996825	DD0280	MBL	normal	M	0	0	0	0	0.49313
GSM996826	DF0319	CLL	tri12	U	1	1	0	0	0.74059
GSM996827	DN0063	CLL	del13	U	1	1	1	NA	0.45659
GSM996828	DR0312	CLL	normal	M	0	0	0	NA	0.43168
GSM996829	DT0300	CLL	del13	M	0	1	0	0	0.40331
GSM996830	FA0294	CLL	normal	M	0	1	0	0	0.49240
GSM996831	FA0429	CLL	del11	U	1	1	0	0	0.32457
GSM996832	FB0332	CLL	del13	M	0	1	0	0	0.48709
GSM996833	FC0075	CLL	del13	M	0	0	0	0	0.34612
GSM996834	FD0404	MBL	del17	U	1	1	1	NA	0.46359
GSM996835	FI0152	CLL	del13	M	0	0	0	0	0.14521
GSM996836	FM0309	CLL	normal	M	1	1	0	0	0.49383
GSM996837	FP0153	CLL	normal	M	0	0	0	0	0.36508
GSM996838	FR0283	CLL	normal	U	1	1	1	0	0.34487
GSM996839	FS0155	MBL	tri12	U	1	1	0	0	0.37417
GSM996840	GA0141	MBL	del13	M	0	0	0	NA	0.28567
GSM996841	GA0398	CLL	normal	M	0	1	0	0	0.34574
GSM996842	GB0293	CLL	del13	U	0	1	0	0	0.49362
GSM996843	GC0015	MBL	tri12	U	0	1	0	0	0.68789
GSM996844	GC0147	MBL	tri12	U	0	1	0	0	0.58707
GSM996845	GC0328	CLL	del13	M	0	1	0	0	0.62208
GSM996846	GD0051	CLL	del13	M	0	0	0	0	0.57301
GSM996847	GD0158	CLL	del13	M	0	1	0	NA	0.31946
GSM996848	GD0262	CLL	del13	M	0	0	NA	NA	0.56454
GSM996849	GE0125	MBL	normal	M	0	1	NA	NA	0.50112
GSM996850	GF0281	MBL	tri12	U	0	1	0	0	0.73213
GSM996851	GG0027	CLL	del13	U	1	1	1	0	0.41445
GSM996852	GL0368	MBL	normal	M	0	1	0	1	0.00000
GSM996853	GN0095	CLL	del13	M	0	0	0	0	0.13506
GSM996854	GP0171	CLL	del11	U	0	1	0	0	0.64181
GSM996855	GP0278	CLL	normal	M	0	1	0	0	0.21128

GEO ID	Patient ID	Disease	Cytogenetics	IGHV	CD38	ZAP70	NOTCH1	SF3B1	Tri12 Score
GSM996856	GP0348	CLL	del13	M	0	0	0	0	0.44339
GSM996857	GR0092	CLL	del13	U	0	1	0	0	0.37098
GSM996858	GR0169	MBL	tri12	M	1	0	0	0	0.46465
GSM996859	GS0343	MBL	normal	U	0	0	NA	NA	0.45829
GSM996860	HG0135	CLL	del13	U	0	0	0	NA	0.52136
GSM996861	IA0025	CLL	normal	M	0	1	0	0	0.44776
GSM996862	IL0066	CLL	del13	U	0	1	NA	NA	0.28039
GSM996863	LF0102	MBL	del13	M	0	0	1	0	0.33541
GSM996864	LI0133	CLL	del13	M	0	0	0	NA	0.14345
GSM996865	LR0247	CLL	normal	M	0	1	0	0	0.30846
GSM996866	LV0118	CLL	normal	U	1	1	1	0	0.31338
GSM996867	MA0088	MBL	del13	M	0	1	NA	NA	0.59122
GSM996868	MA0151	CLL	del13	M	0	0	0	NA	0.09986
GSM996869	MC0055	MBL	tri12	U	1	1	1	0	1.00000
GSM996870	MC0105	CLL	del13	U	0	1	0	0	0.42972
GSM996871	MF0048	MBL	normal	M	0	1	0	0	0.45585
GSM996872	MF0274	CLL	del13	M	0	0	0	0	0.43271
GSM996873	MF0379	CLL	normal	U	0	1	0	0	0.19822
GSM996874	MG0110	CLL	normal	M	0	0	1	NA	0.12854
GSM996875	MG0248	CLL	del13	M	0	1	NA	0	0.17626
GSM996876	MG0449	CLL	normal	U	1	1	1	0	0.26929
GSM996877	MM0097	MBL	normal	U	0	1	NA	NA	0.45196
GSM996878	MM0330	CLL	normal	M	0	0	0	0	0.50210
GSM996879	MN0167	CLL	normal	M	0	0	0	0	0.36572
GSM996880	MO0174	CLL	del13	M	0	0	0	0	0.21070
GSM996881	MO0323	CLL	normal	M	0	1	0	NA	0.49364
GSM996882	MS0115	CLL	normal	U	1	1	1	NA	0.28650
GSM996883	MS0189	CLL	normal	M	0	0	0	NA	0.38264
GSM996884	MS0321	MBL	del13	M	0	1	0	0	0.40600
GSM996885	MS0357	CLL	normal	M	0	0	0	0	0.27935
GSM996886	MV0334	CLL	del13	U	1	1	1	1	0.48030
GSM996887	NB0333	MBL	del13	M	0	1	0	0	0.74216
GSM996888	NG0068	MBL	tri12	U	1	1	1	0	0.41519
GSM996889	NM0156	CLL	normal	U	1	0	0	0	0.05164
GSM996890	NS0060	CLL	normal	U	0	0	0	0	0.20843
GSM996891	OI0282	CLL	normal	M	0	1	0	0	0.4224
GSM996892	OM0201	CLL	del13	M	0	1	0	0	0.45758
GSM996893	OV0127	CLL	normal	M	0	1	NA	NA	0.22170
GSM996894	PA0145	MBL	normal	U	0	1	0	0	0.44651
GSM996895	PA0146	MBL	del13	M	0	0	0	0	0.42542
GSM996896	PA0234	CLL	del13	M	0	0	0	0	0.61943
GSM996897	PA0271	CLL	del13	M	0	0	0	0	0.29204
GSM996898	PA0351	CLL	del13	M	0	0	0	0	0.58424
GSM996899	PC0304	MBL	tri12	U	1	1	0	0	0.69662
GSM996900	PD0164	CLL	del13	M	0	0	0	0	0.39776
GSM996901	PF0177	MBL	normal	U	1	0	0	0	0.30598
GSM996902	PG0028	CLL	del17	U	1	1	0	0	0.47814
GSM996903	PG0216	MBL	normal	NA	0	1	1	0	0.32396

GEO ID	Patient ID	Disease	Cytogenetics	IGHV	CD38	ZAP70	NOTCH1	SF3B1	Tri12 Score
GSM996904	PM0117	CLL	normal	M	0	1	0	0	0.27185
GSM996905	PM0340	MBL	del13	M	0	0	0	0	0.59481
GSM996906	PP0195	MBL	normal	U	1	1	1	0	0.25213
GSM996907	PS0212	CLL	normal	M	0	0	0	0	0.42616
GSM996908	PV0286	CLL	normal	U	0	1	0	0	0.04232
GSM996909	PV0322	CLL	del11	U	1	1	0	0	0.40200
GSM996910	RA0023	MBL	tri12	U	1	1	1	1	0.50547
GSM996911	RA0211	CLL	normal	M	1	0	0	0	0.41832
GSM996912	RA0288	CLL	normal	M	0	1	0	NA	0.50227
GSM996913	RB0040	CLL	normal	M	1	0	NA	NA	0.42300
GSM996914	RD0296	CLL	del13	U	1	1	1	0	0.15405
GSM996915	RG0077	CLL	del13	M	0	1	0	0	0.24435
GSM996916	RG0443	CLL	del11	U	1	1	0	1	0.31199
GSM996917	RM0400	CLL	normal	M	0	0	0	0	0.66530
GSM996918	RR0157	CLL	del13	M	0	0	0	NA	0.14301
GSM996919	RS0260	CLL	normal	U	0	1	0	0	0.34611
GSM996920	RS0325	CLL	del11	U	1	1	0	0	0.27408
GSM996921	SA0021	MBL	del13	M	0	0	0	0	0.58041
GSM996922	SA0093	CLL	del11	U	0	1	0	NA	0.49084
GSM996923	SA0205	MBL	normal	M	0	0	1	0	0.10241
GSM996924	SA0219	MBL	del13	M	0	0	NA	NA	0.49275
GSM996925	SF0098	CLL	del13	M	0	0	0	0	0.43085
GSM996926	SG0114	CLL	normal	U	1	1	0	NA	0.45538
GSM996927	SG0168	MBL	del13	M	0	0	0	0	0.56400
GSM996928	SL0083	CLL	normal	U	1	1	NA	NA	0.51003
GSM996929	SS0299	CLL	normal	M	0	1	0	NA	0.45690
GSM996930	TG0139	CLL	del11	U	0	1	0	0	0.52616
GSM996931	TL0397	CLL	del13	M	0	1	0	0	0.39970
GSM996932	TM0259	MBL	tri12	M	0	1	NA	NA	0.40300
GSM996933	UL0266	CLL	normal	M	0	1	0	0	0.20480
GSM996934	VB0013	CLL	del13	M	0	0	0	0	0.23172
GSM996935	VC0142	CLL	del13	M	0	0	0	NA	0.43119
GSM996936	VC0315	CLL	normal	U	1	1	0	0	0.46466
GSM996937	VD0026	MBL	del13	M	1	1	0	0	0.38496
GSM996938	VD0354	MBL	del13	M	0	1	0	0	0.31554
GSM996939	VG00442	CLL	del13	U	1	1	0	NA	0.37768
GSM996940	VP0089	MBL	del13	M	0	0	0	0	0.62468
GSM996941	VS0016	MBL	normal	M	1	1	0	0	0.27105
GSM996942	VS0140	CLL	normal	M	0	0	0	NA	0.56084
GSM996943	WB0269	MBL	normal	U	0	1	0	0	0.29992

Table S3. Tri12 up-regulated genes identified from isogenic-controlled tri12 transcriptome

Gene products located on chromosome 12, indicated by shading.

Gene Symbol	Fold Change	FDR P-value
DDX3Y	14.27	1.41E-10
RPS4Y1	9.94	6.40E-10
EIF1AY	8.53	5.31E-08
TMPRSS11E	7.98	7.62E-05
USP9Y	7.68	4.50E-06
NLGN4Y	6.77	2.17E-07
UTY	6.77	8.05E-08
KDM5D	5.36	8.84E-06
EDNRB	4.22	4.21E-05
TBX5	3.60	3.89E-07
SULT1C4	3.43	5.38E-05
CEP85L	3.37	6.76E-06
PCDHB12	2.97	1.04E-06
NME8	2.93	1.23E-06
RNASE1	2.91	6.59E-04
KITLG	2.79	2.51E-04
NNAT	2.65	1.83E-04
PTPRB	2.56	3.19E-06
FAM198B	2.51	1.52E-03
PCDHB16	2.50	1.84E-03
CHMP4C	2.48	1.02E-04
ARRB1	2.43	5.73E-05
LCP1	2.43	1.91E-03
IRAK4	2.42	1.46E-05
PUS7L	2.39	2.72E-04
NTS	2.38	1.26E-03
PCDH10	2.37	1.74E-03
TFF1	2.33	4.75E-04
APOA1	2.31	9.51E-04
ALX1	2.29	7.86E-06
GMPR	2.27	1.15E-05
SP140L	2.22	2.62E-04
ZNF257	2.19	8.46E-04
NLRP2	2.14	4.82E-06
BMP4	2.13	2.47E-04
ZNF880	2.12	4.80E-06
SLC39A8	2.11	4.38E-05
SLCO1A2	2.10	2.83E-05
TCP11L2	2.09	2.75E-05
GNA14	2.07	6.47E-04
PCDHB9	2.04	4.95E-08
FOLR1	2.03	1.37E-04
FN1	2.01	2.15E-03

Table S4. Tri12 Score Index.

Due to extremely large file size, not included in thesis.

Table S5. Tri12 up-regulated target genes in CLL reference epigenome.

Gene products with commercially available chemical inhibitors indicated by shading.

ABLIM2	CHAD	GPD1L	LIX1	OTULIN	RRBP1	SYNPO	ZBTB16
AC022431.2	CRYBG3	GTDC1	LRIG1	PDZD2	RUFY3	TAF4B	ZBTB18
ANKHD1-EIF4EBP3	DIRAS1	HIST3H2A	MAP3K1	PHLDB2	SLC12A2	TBC1D9	ZMAT3
ANTXR2	E2F2	IL17REL	MCM9	PHLPP2	SLC20A1	TLL2	ZNF652
AVEN	EML6	KDM1B	MOXD1	PIP5K1B	SLC44A1	TSPAN18	
BACH2	EVI5	KDM4A	MPRIP	PODXL	SMAD3	TWSG1	
C1orf228	FAM160B1	KIAA1841	MREG	RAD21	SNX2	UBE2O	
C8orf88	FHDC1	KIF2C	MTRR	RBM17	SPAG1	UGGT1	
CDK19	GNB4	LBR	NCK1	RIOK3	SSBP3	VCL	

Table S6. Statistical summaries of chemical screening.

*For all statistical tests, one-tailed T tests at 99% confidence interval (alpha = 0.01) were completed if tri12 log GI50 was lower than respective control, in order to compare log GI50 in instances of zero variance. Type 1 error = 5%, due to 5 tests in each group. CLL drug testing completed with CLL#1 and CLL#2. *Abbreviations*: Bruton's tyrosine kinase, BTK; degrees of freedom, df; IRAK(1/)4 inhibitor, IRAK(1/)4i; standard error of log GI50, SE.

		hPSC drug testing						
		tri12 hPSC			**isogenic hPSC**			T-Test P-value
Target	Inhibitor	logGI50	SE	df	logGI50	SE	df	
BTK	ibrutinib	3.490	0.070	20	3.190	0.050	40	
Tri12 gene targets identified in reference epigenome								
CDK19	ponatinib	4.616	0.655	20	4.023	0.172	40	
	CCT251545	4.303	0.498	20	3.394	0.113	40	
SLC12A2	bumetanide	4.000	0.000	20	4.000	0.000	20	
	furosemide	4.000	0.000	20	4.000	0.000	20	
SLC44A1	hemicholinium-3	4.303	0.348	20	2.765	0.06	40	
Tri12 gene targets identified using tri12 hPSC platform								
EDNRB	bosentan	3.574	0.076	40	4.095	0.146	30	0.0006
	ambrisentan	4.000	0.000	40	4.000	0.000	40	
IRAK4	IRAK4i	2.263	0.108	50	2.945	0.079	30	< 0.0001
	IRAK1/4i	4.000	0.000	40	4.000	0.000	40	
KIT	axitinib	3.487	0.038	40	4.086	0.059	40	0.0005
	dasatinib	4.000	0.000	20	4.000	0.000	40	
	sunitinib	3.787	0.038	40	4.086	0.059	40	< 0.0001
FOLR1	methotrexate	5.145	0.022	40	5.277	0.037	40	0.0015
		CLL drug testing						
		tri12 CLL			**normal CLL**			T-Test P-value
Target	Inhibitor	logGI50	SE	df	logGI50	SE	df	
BTK	ibrutinib	3.280	0.050	20	3.230	0.070	20	
Tri12 gene targets identified in reference epigenome								
CDK19	ponatinib	2.747	0.135	40	2.535	0.103	40	
	CCT251545	3.516	0.033	20	3.576	0.111	20	
SLC12A2	bumetanide	4.000	0.000	50	4.000	0.000	30	
	furosemide	4.000	0.000	50	4.000	0.000	30	
SLC44A1	hemicholinium-3	3.804	0.083	50	3.772	0.042	50	
Tri12 gene targets identified using tri12 hPSC platform								
EDNRB	bosentan	3.539	0.048	40	4.710	0.106	40	< 0.0001
	ambrisentan	4.000	0.000	50	3.519	0.040	30	
IRAK4	IRAK4i	2.612	0.037	50	3.056	0.113	50	0.0002
	IRAK1/4i	2.863	0.068	40	3.405	0.290	40	0.0363
KIT	axitinib	3.687	0.099	20	4.000	0.000	20	0.0020
	dasatinib	3.590	0.082	40	3.427	0.082	20	
	sunitinib	2.997	0.036	40	2.826	0.079	20	
FOLR1	methotrexate	4.679	0.106	50	5.085	0.055	30	0.0028

Table S7. Primary use of drugs identified.

(‡) Diseases of the hematopoietic and immune system indicated by grey shading. (*) Multi-targeted receptor tyrosine kinase inhibitor (TKI). *Abbreviations*: endothelin receptor A, EDNRA; IRAK(1/)4 inhibitor, IRAK(1/)4i; platelet-derived growth factor receptor family, PDGFR; and vascular endothelial growth factor receptor family, VEGFRs.

Drug	Disease(s) ‡	Putative Target(s)	Trial
Bosentan	pulmonary arterial hypertension	EDNRA, EDNRB	NCT00091715
Ambrisentan	pulmonary arterial hypertension	EDNRA	NCT00091598
Axitinib *	renal cell carcinoma, vestibular schwannomas, head and neck cancer	PDGFR (α, β), VEGFR (1-3), KIT	NCT03595124, NCT02762513
	chronic myeloid leukemia	BCR-ABL(T315I)	Pemovska et al., 2015
Dasatinib *	chronic myeloid leukemia, and acute lymphoblastic leukemia	BCR-ABL	NCT00481247, NCT00391989
Sunitinib *	renal cell carcinoma, imatinib-resistant gastrointestinal stromal tumor	PDGFR (α, β), VEGFR (1-3), KIT	NCT00930033
IRAK4i	non-Hodgkin lymphoma	IRAK4	NCT03328078
IRAK1/4i	T cell acute lymphoblastic leukemia	IRAK1, IRAK4	Li et al., 2015
	myelodysplastic syndrome	IRAK1, IRAK4	Rhyasen et al., 2013

Chapter 3: CXCL12/CXCR4 signaling enhances human PSC-derived hematopoietic progenitor function and overcomes early *in vivo* transplantation failure

Preamble

This chapter is an original published article. It is presented in its published form.

DOI: https://doi.org/10.1016/j.stemcr.2018.04.003

Author contributions: J.C.R., B.T., D.G., A.L.B., and D.P.P. performed experiments. T.J.C. provided technical assistance for the calcium flux assay. J.C.R. and M.B. designed experiments, interpreted data, and wrote the manuscript. M.B. directed the study.

Author contributions in greater detail: J.C.R. wrote the manuscript and performed all experiments except for the following: B.T. cloned wildtype CXCR4 into a lentiviral vector, produced 50% of the CXCR4 lentivirus used in this study, and provided feedback on the manuscript. In 25% of *in vitro* experiments, D.G. prepared media, buffers, and antibody mixes for flow cytometry; cryopreserved cells; thawed cells; and provided feedback on the manuscript. A.L.B. provided feedback on my statistical analyses, performed flow cytometry of cells used for transplantation, and provided feedback on the manuscript. D.P.P. plated CFU for one experiment. T.J.C. changed microscope filters and wrote a protocol for the calcium flux assay. M.B. designed experiments, interpreted data, directed the study, and provided feedback on the manuscript.

CXCL12/CXCR4 signaling enhances human PSC-derived hematopoietic progenitor function and overcomes early *in vivo* transplantation failure

Jennifer C. Reid,[1,2] Borko Tanasijevic,[1,3] Diana Golubeva,[1,2,3] Allison L. Boyd,[1] Deanna P. Porras,[1,2] Tony J. Collins,[1] and Mickie Bhatia[1,2,*]

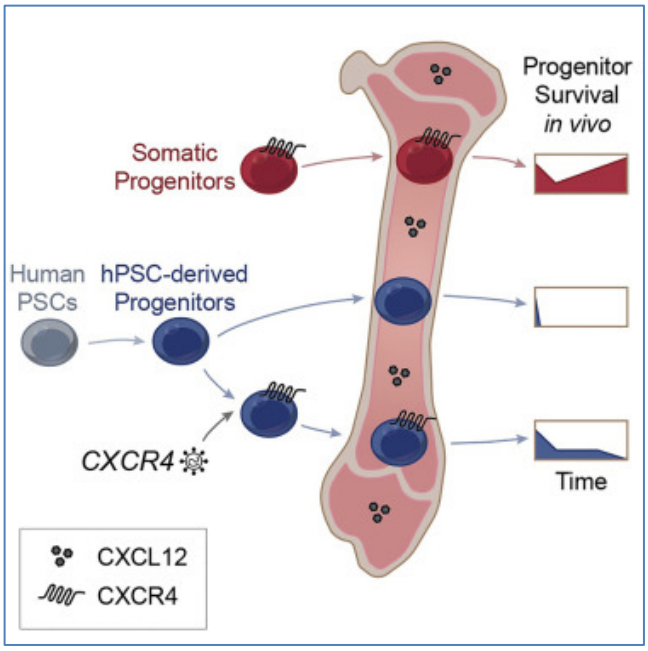
[1]Stem Cell and Cancer Research Institute, Michael G. DeGroote School of Medicine, McMaster University, Hamilton, ON L8N 3Z5, Canada
[2]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON L8N 3Z5, Canada
[3]These authors contributed equally
*Correspondence: mbhatia@mcmaster.ca (M.B.)

Running Title: CXCR4 signaling in hematopoietic progenitors from human PSCs

Journal cover image

***Stem Cell Reports*, Volume 10, Issue 5**: Reid et al. (pp. 1625–1641) report that forced expression of CXCR4 by human pluripotent stem cell-derived hematopoietic progenitors enables directed navigation *in vitro* and enhances survival following transplantation in the bone marrow *in vivo*, where these cells otherwise fail without CXCR4. The cover is an artistic illustration on the theme, where a human can better navigate and survive in the ocean upon the addition of scuba gear. In contrast, humans can only transiently stay in the ocean without scuba gear. Cover created by Jennifer Reid.

Graphical Abstract



Highlights

- Transplant kinetics indicate human PSC-HPCs fail to survive within the first 24 h in bone marrow
- hPSC-HPCs aberrantly express chemokine receptors, as they lack CXCR4 but express CX3CR1
- Ectopic CXCR4 signaling enhances hPSC-HPC function *in vitro* and upon transplantation *in vivo*
- CXCR4 exerts global transcriptional shift of CXCR4^{+} hPSC-HPCs towards adult/somatic HPCs

eTOC

Bhatia and colleagues reveal that human PSC-derived hematopoietic progenitor cells fail to survive even 24 hours following *in vivo* bone marrow transplantation, while these same progenitors survive and proliferate for weeks *in vitro*. They link these observations to deficiencies in CXCR4 signaling, which when rectified, leads to enhanced progenitor function and survival in the bone marrow, and a transcriptional shift towards somatic hematopoietic stem cells gene profiles.

Summary

Human pluripotent stem cells (hPSC) generate hematopoietic progenitor cells (HPC), but fail to engraft xenograft models used to detect adult/somatic hematopoietic stem cells (HSC) from donors. Recent progress to derive hPSC-derived HSCs has relied on cell autonomous

forced expression of transcription factors, however the relationship of bone marrow (BM) to transplanted cells remains unknown. Here, we quantified a failure of hPSC-HPCs to survive even 24 h post-transplantation. Across several hPSC-HPC differentiation methodologies, we identified the lack of CXCR4 expression and function. Ectopic CXCR4 conferred CXCL12 ligand-dependent signaling of hPSC-HPCs in biochemical assays and increased migration/chemotaxis, hematopoietic progenitor capacity, as well as survival and proliferation following *in vivo* transplantation. This was accompanied by a transcriptional shift of hPSC-HPCs toward somatic/adult sources, but this approach failed to produce long-term HSC xenograft reconstitution. Our results reveal that networks involving CXCR4 should be targeted to generate putative HSCs with *in vivo* function from hPSCs.

Introduction

Hematopoietic stem cell (HSC) transplants are the only globally adopted stem cell therapy for patients, and have been shown to be curative for hematological malignancies and diseases along with certain solid tumors (Copeland, 2006). However, given the scarcity of compatible donors against the number of patients in need (Gratwohl et al., 2015), developing alternative sources of HSCs is paramount. While hematopoietic progenitors (HPC) can readily be generated by human pluripotent stem cells (hPSCs*) in vitro*, they lack robust engraftment potential (Gori et al., 2015; Lee et al., 2017a; Ng et al., 2016; Risueño et al., 2012; Wang, 2005). Ectopic TF expression has been used in attempts to induce BM engraftment of hPSC-HPCs (Doulatov S et al., 2017; Doulatov et al., 2013; Ramos-Mejia et al., 2014; Ran et al., 2013; Wang, 2005). Recently, expression of 7 TFs in hPSC-derived hemogenic endothelium generated HSC/HPCs, but only after BM transplantation (Sugimura et al., 2017). Despite this progress, the *in vivo* dependent approach did not produce an abundance of HSCs and these cells remain molecularly unrelated to somatic HSCs; factors which require modification for successful clinical translation.

Unlike solid organ transplants, injected HSCs must migrate to and reside in specialized niches in the BM, the primary site of adult hematopoiesis (Boyd and Bhatia, 2014). Adult HSCs receive complex and dynamic cues from the BM for survival, quiescence, homeostasis, and proliferation. Likewise, using co-cultures of BM stroma cells or embryonic niche cells improves hPSC-HPC derivation (Ledran et al., 2008; Tian et al., 2006; Vodyanik et al., 2006; Weisel et al., 2006), suggesting these cells too, require specific niche cues. BM secreted CXCL12 (formerly known as SDF1) is a powerful regulator of HSC function and binds its cognate receptor, CXCR4, expressed by HSC/HPCs (Lapidot and Kollet, 2002; Nagasawa et al., 1996; Sugiyama et al., 2006). CXCR4 represents the sole chemokine receptor utilized by HSCs for migration/chemotaxis (Wright et al., 2002) and regulates the proliferation of somatic HSCs (Kahn et al., 2004). This is sustained through an auto-regulatory loop that is dynamically regulated from cell surface to intracellular stores (Lapidot and Kollet, 2002). CXCR4 is regulated by BM factors, some of which include hypoxia (Scheurer et al., 2004), and Notch (Wang et al., 2017), glucocorticoid (Guo et al., 2017), and prostaglandin E_2 (Goessling et al., 2011) signaling pathways. However, the functional capacity of hPSC-HPCs to respond to BM regulatory cues remains largely unknown.

Previous studies assessing hPSC-HPC engraftment potential have reported low levels of human hematopoietic microchimerism in immunocompromised mouse BM 4 or more weeks post-transplant (Doulatov S et al., 2017; Doulatov et al., 2013; Gori et al., 2015; Lee et

al., 2017a; Ramos-Mejia et al., 2014; Ran et al., 2013; Risueño et al., 2012; Wang, 2005). Here, we reveal the previously unappreciated early transplantation failure of hPSC-HPCs *in vivo* that occurs within the first 24 hours, despite robust hematopoietic progenitor capacity detected for weeks *in vitro*. Across a broad range of differentiation methodologies, global transcriptional analysis identified the chemokine receptor CXCR4 as aberrantly regulated in hPSC-HPCs. Molecular and biochemical characterization of ectopic CXCR4 in hPSC-HPCs demonstrated that CXCR4 enhances survival and progenitor retention following BM transplantation. We propose the CXCR4 network is essential for physiological responsiveness towards generating bonafide HSCs from hPSCs.

Results

Defective in vivo *retention of hPSC-HPCs*

Early properties of hPSC-HPC integration into the BM *in vivo* have not been explored by direct side by side comparisons to human adult/somatic HPC sources. Cord blood (CB) is readily available for experimentation as a somatic source of HSCs that establish long term multilineage hematopoietic engraftment in xenograft models (Boyd et al., 2017). Furthermore, transplantation of CB cells has been used clinically for long term reconstitution of donor derived healthy hematopoiesis in patients (Cutler et al., 2013). As such, we used CB as a source of transplantable cells to analyze early HPC behavior and compare this directly to HPCs derived from hPSCs for the first time. hPSC-derived HPCs were derived using embryoid body (EB) formation and differentiated with hematopoietic cytokines and BMP4 (Chadwick et al., 2003), and were utilized on EB day 15 for analysis and transplantation. Somatic and hPSC-HPCs do not share equivalent frequencies of phenotypic or functional progenitors, as quantified by human specific $CD34^+CD45^+$ cell surface expression (vs. mouse mCD45; Figure 1A) and colony forming unit (CFU) composition (Figure S1A), respectively. These results are consistent with previous reports across a broad range of methodologies to produce phenotypic or functional progenitors from hPSCs (Doulatov et al., 2013; Lee et al., 2017a; Ramos-Mejia et al., 2014; Risueño et al., 2012; Saxena et al., 2016; Tian et al., 2006; Vodyanik et al., 2006), as well as non-human primate *Macaca nemestrina*-induced PSC-derived HPCs (Gori et al., 2015). To accurately control for the number of input human $CD34^+CD45^+$ cells and functional progenitors injected from hPSC-HPC cultures, we therefore selected two doses of CB for transplantation (Figure 1B). The higher CB dose, CB^{CD34}, closely approximates the number of total phenotypic hPSC-HPCs transplanted (dark red, $5x10^4$ $CD34^+$ cells). The lower dose, CB^{PRO}, closely approximates the total number of functional hPSC-HPC progenitors, as well as the clinical dose of $2x10^5$ $CD34^+$ cells/kg (Gluckman, 2009) when scaled by mass to a 25 gram mouse (light red, $5x10^3$ $CD34^+$ cells). Total phenotypic and functional HPCs prior to injection were measured at Day 0 (Figure 1A and 1B).

Using this carefully quantitated approach to phenotypically and functionally enumerate equivalency of transplanted cells, human CB vs. hPSC-derived HPCs were injected into the femurs of murine recipients, where the BM was assessed for human chimerism at the functional and phenotypic level at multiple time points within the first week. At the same time points as injected femur assessment, we determined migration capacity *in vivo* by analysis of contralateral femur BM, spleen, and lungs (Figure 1C). The number of individual mice from 4 transplant groups were compared at 24 h and 2, 3, and 5 days as indicated

(Figure 1D). The frequency of human hematopoietic cell chimerism was rare at these early times, but could be captured by flow cytometric analysis for human HPCs ($mCD45^-$ $hCD45^+CD34^+$, Figure 1E, Figure S1B and S1C). Phenotypic CB HPC expansion was evident within the injected femur BM well within this time frame (Figure 1E). As predicated (Wang, 2005), intra-femoral injection provided an engraftment advantage to retain HPCs in the injected femur, while a subpopulation of somatic HPCs could still home to distal BM sites such as the contralateral femur (Figure 1F), but not to extramedullary sites such as the lung (Figure S1D-S1F), or spleen (Figure S1G-S1I). In contrast, hPSC-HPCs were not able to persist even at 24 h post-transplant in any location, yet CB HPCs were capable of robust and exponential expansion in the BM at both cellular doses (Figure 1G and 1H; comparison of fits test, Prism software). Due to the rarity of these cells at early time points post-transplantation, evaluation of HSC function by secondary transplantation was not feasible for both somatic and hPSC-derived HPCs. However, we were able to extract BM and compare early engraftment kinetics at the progenitor level using the CFU assay *in vitro*. Functional CB HPCs were retained in injected femur BM and continued progenitor output over the first week, while putative hPSC-HPCs failed to be retained as measured by erythroid-myeloid CFU assays (Figure 1J-1L). Strikingly, BM retention of CB HPCs was strongly correlated between phenotypic and functional measures, whereas no such relationship existed for hPSC-HPCs devoid of functional progenitors (Figure 1M). These results suggest a reduction of somatic HPCs within the first 24 h post-injection, followed by a rapid increase from 24 h to 5 days *in vivo*. This behavior is in sharp contrast to hPSC-HPCs, that have both phenotypic expression and functional capacity similar to CB cells when transplanted into immune deficient mice, but fail to recover and proliferate *in vivo*. These experimental observations reveal an unappreciated deficiency contributing to hPSC-HPC engraftment failure that occurs upon initial transplantation.

Somatic and hPSC-derived HPCs are functionally similar in vitro

Based on the inability of hPSC-HPCs to survive and proliferate *in vivo*, we examined their behavior *in vitro* across a broad range of hPSC differentiation methodologies, and compared them relative to adult/somatic control sources of HPCs. Somatic HPCs can be harvested from human BM, adult mobilized peripheral blood (MPB), and neonatal CB sources, and are enriched in the $CD34^+CD45^+$ subpopulation (Figure 2A). Similarly, the past decade has provided several methodologies to derive hPSC-HPCs, and so we investigated three very different protocols and approaches, 1) Cytokines and BMP4 treatment of embryoid bodies (EB) (Chadwick et al., 2003), 2) OP9 co-culture (Vodyanik et al., 2006), and 3) endothelial-hematopoietic transition (EHT) (Lee et al., 2017a) (Figure 2B). We routinely observed hPSC-derived hematopoietic CFU morphology similar to somatic HPCs when cultured for 14 d *in vitro* (Figure 2C and 2D). To control for variations in $CD34^+CD45^+$ frequency, we determined the CFU output and lineage distribution per 1000 $CD34^+CD45^+$ cells, which was similar across adult BM, adult MPB, and neonatal CB donor samples (Figure 2E). Similarly, while the three differentiation methods produced varying frequencies of hPSC-HPCs (Figure 2B), CFU per 1000 HPCs was also similar (Figure 2F), except for adherent EHT-derived cells which are reportedly less enriched for progenitors than non-adherent cells (Lee et al., 2017a). hPSC-HPCs readily survive and proliferate *in vitro* (Figure 2), and yet failed to survive very short time frames *in vivo* (Figure 1). Even the same hPSC-HPCs tested in parallel in these two

environments (*in vivo* versus *in vitro*) lead to sharply contrasting effects in progenitor capacity (Figure 1B and 1K). These comparative results indicate hPSC-HPCs may be insufficiently responding to BM environmental cues *in vivo* that prevent the survival that can be readily demonstrated *in vitro*.

Deficient chemokine receptor expression is a consistent feature of hPSC-HPCs

To understand potential underlying interactions that differ between hPSC-HPCs and BM versus adult/somatic HPCs and BM that successfully engraft in vivo, we applied global transcriptome analysis of primitive hematopoietic populations enriched for HSC/HPCs (CD45+CD34+CD38–), from fifteen healthy donors including BM, MPB, CB, and fetal blood (FB) sources (Figure 3A, see also Supplemental Experimental Procedures) and collectively compared these to hPSC-HPCs derived samples using the Cytokines and BMP4 (Chadwick et al., 2003), and EHT method of differentiation (Lee et al., 2017a), as well as using gene expression data for TF-expressing hPSC-HPCs (Doulatov et al., 2013). Gene set enrichment analysis (GSEA) of global transcriptional profiles identified several gene sets relevant to environmental cues in vivo, including chemokine receptors, extracellular matrix interactions, integrins, and cell surface interactions at the vascular wall (Figure 3B, Table S1), which were enriched in hPSC-HPCs. Closer observation identified two consistent differentially expressed chemokine receptor genes, *CXCR4* and *CX3CR1* (Figure 3C), which are associated with BM retention and egress, respectively (Nakano et al., 2017). We then performed global transcriptome analysis comparing the four different sources of hPSC-HPCs individually to somatic HPCs (MPB, BM, CB, and FB), which identified 666 consistently differentially expressed genes (Figure 3D, Table S2). Unsupervised hierarchical clustering segregated 120 as exclusively expressed by somatic HPCs (Figure 3E). Within these 120 genes, we observed *HOXA5*, *HOXA10*, *FLT3,* and *PROM1*, and epigenetic regulators such as *HDAC7*, *KAT6A*, and *MLLT3*, as well as *CXCR4* (Table S2). The STRING protein-protein interaction database (Szklarczyk et al., 2015) identified potential interactions of the 120 HSC-related genes, where CXCR4 is prominently linked to these key HSC and HPC-related genes (Figure 3F).

These consistent observations suggested a potentially broad deficiency that we further investigated with flow cytometric analysis, which revealed that CXCR4 and CX3CR1 proteins are mutually exclusively expressed on the cell surface (Figure S2A-S3C). Furthermore, CXCR4 is enriched on CB HPCs (Figure S2A), while CX3CR1 is enriched on hPSC-HPCs (Figure S2B and S2C). We expanded upon these observations with three diverse sources each of somatic and hPSC-derived HPCs (Figure 3G-3J). We observed a highly reproducible $CXCR4^{+}CX3CR1^{-}$ phenotype expressed by somatic HPCs and absence of this subset from hPSC-HPCs, which instead were $CX3CR1^{+}CXCR4^{-}$. Importantly, somatic HPCs engrafted in mice >4 weeks continued to express CXCR4 while being absent for CX3CR1 expression (Figure S2D). Low expression of cutaneous lymphocyte antigen (CLA, HECA452 clone) was recently suggested to have a role in hPSC-HPC transplantation deficiency (Lee et al., 2017a). While we could replicate low CLA expression relative to a somatic HPC source, we again did not observe CXCR4 expression by transcriptional or protein analyses of EHT-derived hPSC-HPCs (Figure S2E and S2F), suggesting both may be important factors. $CXCR4^{-}$ and $CX3CR1^{+}$ phenotypes of hPSC-HPCs have been previously noted (Ng et al., 2016; Salvagiotto et al., 2008; Tian et al., 2006) and reinforce our observations that low CXCR4

expression is a consistent deficiency of hPSC-HPCs across a broad range of methodologies and hPSC cell lines.

To move beyond phenotypic observations, we physically isolated CB and hPSC-derived hematopoietic cells based exclusively on chemokine receptor expression, and determined biological hematopoietic progenitor function. As predicted, hematopoietic progenitors measured by CFU potential were enriched within $CXCR4^+$ CB cells, and not observed from $CX3CR1^+$ cells (Figure 4A-4C), as CX3CR1 expression is restricted to mature hematopoietic cells (Nakano et al., 2017). In stark contrast, CFU potential was observed from $CX3CR1^+$ hPSC-derived hematopoietic cells, while $CXCR4^+$ hPSC-derived cell CFU potential was significantly reduced (Figure 4D-4F). CXCR4 and CX3CR1 are chemokine receptors that bind only one cytokine; CXCL12 and CX3CL1, respectively (Balkwill, 2004). CXCL12 is most highly expressed in BM tissue, whereas CX3CL1 is more highly expressed in non-hematopoietic and extramedullary sites such as the lungs, kidneys, and spleen (Uhlén et al., 2015). We therefore sought to determine whether the CXCR4 or CX3CR1 receptors expressed by hPSC-hematopoietic cells (Figure 3E) were functionally capable of initiating chemotaxis towards their cognate ligands (Figure 4G), to investigate whether hPSC-HPCs may be receiving BM cues for retention or egress. Serving as a positive control, bulk somatic hematopoietic cells were able to migrate towards both CXCL12 and CX3CL1 (Figure 4H), as both receptors were observed on bulk CB cells (Figure 3D). MPB was an exception, which exhibited a blunted response towards CXCL12 (Figure 4H), which is likely correlated to CXCR4-antagonist treatment (AMD3100) in donors to mobilize healthy HSCs and HPCs out of the BM and into circulation (Broxmeyer et al., 2005). Despite multiple experiments, hPSC-HPCs did not migrate towards CXCL12 or CX3CL1 (Figure 4H). *In vivo* BM engraftment requires signals mediating BM retention and limiting egress. Due to the lack of chemotactic response exhibited by hPSC-HPCs (Figure 4H), we suggest the lack of CXCL12-CXCR4 function is preventing BM retention, while CX3CL1-CX3CR1 binding is not actively participating in BM egress (Nakano et al., 2017). Together, these data suggest that promoting CXCR4 expression may be beneficial to improve hPSC-HPC function.

Inducing CXCL12-dependent CXCR4 signaling

On the basis of a beneficial functional relationship between CXCR4 and somatic HPCs (Brenner et al., 2004; Kahn et al., 2004), we next sought to pharmacologically induce CXCR4 expression by hPSC-HPCs. We tested several pharmacological agents reported to upregulate CXCR4, including cyclic AMP (cAMP) agonists, forskolin and prostaglandin E_2 (Goessling et al., 2011; Saxena et al., 2016), and hormones (flonase and estrogen (Guo et al., 2017; Rodriguez-Lara et al., 2017). While these compounds effectively increased the number of $CXCR4^+$ CB HPCs, this was not observed from hPSC-HPCs (Figure 5A and 5B, Figure S3A and S3B). GSEA pathway analysis demonstrated that networks targeted by these compounds are not equally active (Figure 5C), supporting our observation of an inability to upregulate CXCR4 by hPSC-HPCs.

Therefore, lentiviral ectopic CXCR4 expression was developed and functionally validated (Figure 5D). We developed an additional vector, which expressed CXCL12-unresponsive CXCR4 (N123K mutation (Zhang et al., 2002); termed $CXCR4(off)^+$), in order to identify CXCL12-dependent biological effects. A third vector expressed only GFP (vector control). Transduction with our CXCR4 vectors resulted in robust expression of CXCR4,

which could also be indirectly monitored using GFP expression (Figure 5E). Upon transduction, CXCR4$^+$ hPSC-derived hematopoietic cells could robustly transmigrate towards CXCL12 in Transwells *in vitro* (Figure 5F), a classic feature of adult/somatic HPCs (Wright et al., 2002). CXCR4$^+$ and CXCR4(off)$^+$ hPSC-HPCs expressed similar cell surface CX3CR1 compared to control (Figure S3C), and transduced hPSC-HPCs remained unresponsive to CX3CL1 in chemotaxis assays (Figure S3D). Intracellular calcium signalling was induced by CXCL12 in CXCR4$^+$ hPSC-derived hematopoietic cells (Figure 5G), achieving similar frequencies to that observed in CB (Figure 5H). The response to CXCL12 was not observed in CXCR4(off) transduced cells and could be inhibited by pre-treatment with the CXCR4 antagonist AMD3100 (Figure 5G and 5I). hPSC-derived hematopoietic cells could also trigger calcium flux signaling upon CXCL12 treatment, again only when wildtype CXCR4 was expressed (Figure 5G). CXCR4$^+$ hPSC-HPC calcium flux responses were achieved at similar frequencies to CB (Figure 5H), and were inhibited with the CXCR4 antagonist (Figure 5G and 5I) (Broxmeyer et al., 2005), further demonstrating the functional integration of CXCR4 and the ability to pharmacologically regulate this network within hPSC-HPCs. Additionally, CXCL12 treatment resulted in a CXCR4 dependent two-fold increase in progenitor capacity (Figure 5J). The enhancement in CFU potential is in line with a previous report where CXCL12 treatment in MethoCult increased CB progenitor propagation (Broxmeyer et al., 2007). Overall, this data demonstrates that both CXCL12 and CXCR4 are critically involved in hPSC-HPCs, and that lentiviral expression of CXCR4 enables its functional integration into networks supporting biological processes for which hPSC-HPCs were deficient compared to somatic HPC sources.

Enhanced BM progenitor retention of CXCR4+ hPSC-HPCs

Ectopic CXCR4 has previously been shown to enhance CB engraftment (Kahn et al., 2004). We therefore transplanted GFP tagged CXCR4$^+$ hPSC-HPCs *in vivo* and assessed survival and proliferation. Equivalent numbers of hPSC-HPCs transduced with CXCR4 or vector control were injected into large groups of mice ($n \geq 20$), in parallel with somatic HPCs (Figure 6A and 6B). Encouragingly, CXCR4$^+$ hPSC-HPCs exhibited enhanced BM retention at the fifth day post-transplant (Figure 6C), and lead to a significant increase in phenotypic hPSC-HPC retention overall (Figure 6D, Figure S4A) and also enabled CXCR4$^+$ hPSC-derived HPCs to migrate to the contralateral femur BM similarly to somatic HPCs (Figure S4B). This finding was specific to the BM, as we did not observe HPCs in the spleen (Figure S4C-S4E). Highest frequencies of GFP were observed from the BM of femurs injected with CXCR4$^+$ hPSC-HPCs (Figure S4F). The retention of phenotypic hPSC-HPCs was also paralleled with retention of functional progenitors. Impressively, CXCR4 alone was sufficient to rescue progenitor function following transplantation (Figure 6E and 6F, Figure S4G and S4H). CFU were manually picked and validated for human origin and lentivirus-based *CXCR4* and *GFP* sequences by PCR (Figure 6G and 6H). We ascertained a loss of erythroid progenitors from CXCR4$^+$ hPSC-HPCs *in vivo* (Figure S4G), consistent with previous reports of HPCs derived from non-human primate *Macaca nemestrina* induced PSCs (Gori et al., 2015). CFU retention, measured as a fraction of CFU injected, was equivalent between somatic and CXCR4$^+$ hPSC-derived HPCs, and significantly higher than control hPSC-HPCs (Figure 6I). Unfortunately, CXCR4 was not sufficient to confer sustained engraftment tested at 4 weeks (Figure 6J-6L). This lack of prolonged hematopoietic reconstitution is possibly due to the silencing of CXCR4

vector (loss of GFP) over time (Figure S4I). Our results suggest the importance of identifying hPSC-HPCs which exhibit dynamic auto-regulation of the CXCR4 pathway in response to extracellular stimuli similar to adult/somatic HPCs.

CXCR4 networks are a feature of somatic HSC/HPCs

To investigate CXCR4 auto-regulation, sustained activation, and network responsiveness, global transcriptome analysis of CXCR4$^+$ and CXCR4$^-$ hPSC-HPCs (CD34$^+$CD45$^+$GFP$^\pm$) were compared to somatic and 7 TF-hPSC-derived HSC/HPCs with long term engraftment capacity, 3 and 5 TF-expressing HPSC-HPCs with little to no engraftment capacity, and related BM/niche cell types (Figure 7A). CXCR4$^+$ hPSC-HPCs clustered as an intermediate between CXCR4$^-$ hPSC-HPCs and somatic HPCs. Furthermore, GSEA identified four independent HSC signatures which were significantly enriched in CXCR4$^+$ versus CXCR4$^-$ hPSC-HPCs, in addition to other cellular adhesion molecules (Figure 7B and 7C). Global transcriptome analysis of CXCR4$^+$ and CXCR4$^-$ hPSC-HPCs revealed genes that may shift hPSC-HPCs closer towards somatic HPC profiles as well as genes which continue to be aberrantly expressed (Figure 7D, Table S3). We observe classical HSC genes, PROM1 and CD34, and adhesion genes, SELL and ICAM2, remaining highly expressed by somatic HPCs as compared to CXCR4$^+$ hPSC-HPCs. Interestingly, CXCR4$^+$ express lower levels of the mobilization-inducing MMP9 gene than CXCR4$^-$ hPSC-HPCs, however, still higher than somatic HPCs and therefore considered an "improved gene". Deeper molecular analysis of the transcriptome was performed using Pearson correlation and scoring pathway expression using normalized enrichment scores (NES) from GSEA analyses and BM HPCs as reference (Figure 7E, see also Supplemental Experimental Procedures). CXCR4$^+$ hPSC-HPCs were most similar to BM HPCs compared to other hPSC-derived cells (Pearson correlation). However, negative NES scores of CXCR4 regulatory pathways suggests ectopic CXCR4 did not induce auto-regulation. In contrast, while 7 TF hPSC-HSC/HPCs are molecularly distinct from somatic HSCs (Pearson correlation of 0.77, similar to original report of >0.7 (Sugimura et al., 2017)), they appear to overexpress CXCR4-related pathways, using the NES score as an indicator. The observation of active CXCR4-related pathways in 7 TF hPSC-HSC/ HPCs (Sugimura et al., 2017) and BM/niche cell types suggests that during *in vivo*-programming, BM may select for hPSC-HSC and HPCs capable of receiving extracellular cues for survival. This reinforces the hypothesis of CXCR4 signalling as necessary, but not sufficient, to produce hPSC-HSC and HPCs. To functionally investigate auto-regulation, we treated CXCR4$^+$ hPSC-HPCs with several CXCR4-inducing agents (Figure 7F). Indeed, pathways upstream of CXCR4, such as PGE_2 and glucocorticoids (flonase) (Goessling et al., 2011; Guo et al., 2017), were unable to synergize with ectopic CXCR4 (Figure 7F-7H). However, forskolin, a cAMP agonist, increased total CXCR4$^+$ HPCs while total cellular yield was equivalent, suggesting incomplete CXCR4 auto-regulation.

Discussion

Our study demonstrates that CXCR4 is necessary for early transplantation survival but not sufficient to confer long-term engraftment. We suggest that the loss of CXCR4 leads hPSC-HPCs to revert to CXCR4$^-$ characteristics instead of auto-regulating CXCR4 *in situ* (Figure 7H). Altogether, CXCL12-mediated CXCR4 signalling by hPSC-HPCs promotes progenitor proliferation, survival, migration *in vitro*, and increased BM retention *in vivo*. Progress

towards hPSC-HSC long-term engraftment has largely been pursued using ectopic TF expression. In contrast, somatic HSCs receive non-cell autonomous signals in the BM directing cell fate (Boyd and Bhatia, 2014; Boyd et al., 2017), therefore we investigated these cues in the context of hPSC-HPCs. Our findings identified hPSC-HPC transplantation deficiencies which complements previous reports of limited engraftment at 4 weeks (Gori et al., 2015; Lee et al., 2017a; Ng et al., 2016; Risueño et al., 2012; Wang, 2005), while at a much earlier time frame within the first week post-transplantation.

Cell autonomous approaches to produce hPSC-HSC/HPCs have made significant progress using ectopic HSC-specific TFs (Doulatov S et al., 2017; Doulatov et al., 2013; Ran et al., 2013; Sugimura et al., 2017). However, all four of these reports required TFs with leukemogenic potential, most prominently *RUNX1* (*AML1*) and *ERG* (Crans and Sakamoto, 2001). Furthermore, the additional effects of *in vivo* BM programming during multiplexed TF expression appear to support higher chimerism frequencies of hPSC-derived hematopoietic cells and multilineage differentiation potential (Sugimura et al., 2017). CXCR4 function is critical for BM retention and HSC function (Lapidot and Kollet, 2002; Nagasawa et al., 1996; Sugiyama et al., 2006), however, similar to TF expression it must be tightly regulated. Gain-of-function mutations in CXCR4 is associated with a rare primary immunodeficiency called WHIM (Warts, Hypogammaglobulinemia, Infections, and Myelokathexis) syndrome (Heusinkveld et al., 2017). Clinical features include abnormal retention of mature hematopoietic cells in the BM of patients, which suggest constitutive knock in or long-term induced expression systems may be deleterious for promoting balanced multi-lineage engraftment from hPSC-HPCs. This points to the importance of producing hPSC-HPCs which can auto-regulate CXCR4 activity, without prolonged ectopic expression altogether.

Our study establishes a proof of principle that supporting CXCR4 expression is a viable target not only to enhance progenitor survival and proliferation when exposed to CXCL12 *in vitro* or BM *in vivo*, but also to promote a more normal hematopoietic transcriptional shift towards somatic HSC and HPCs. Independent of ectopic chemokine receptor modulation, several other lines of evidence provide support for required functional connectivity between BM niche and hPSC-HPCs for sustained engraftment. The injection of undifferentiated human PSCs into mice has led to the rare production hPSC-HSCs isolated from the BM (Suzuki et al., 2013), when the original cells were injected as heterotopic teratomas. Conversely, once isolated from their niche, somatic HSCs are difficult to propagate in vitro, and prolonged culture leads to diminished engraftment potential (Brenner et al., 2004; Hofmeister et al., 2007). Strikingly, the first clinical report of a pharmacological agent able to expand HSCs *in vitro* involved Notch signaling, which operates in a non-cell autonomous role in the BM (Delaney et al., 2010). Together, these relationships underscore the importance of BM cues for hPSC-HPCs as an unappreciated area of biological interaction that requires further investigation for future clinical applications of human HPCs derived from hPSC sources.

Experimental Procedures

Statistics
Data are represented as mean ± SEM. Prism (6.0c, GraphPad) software was used for all statistical analyses, and the criterion for statistical significance was $P < 0.05$. Statistics are described in figure legends.

Data availability
Microarray data from this study have been deposited in the Gene Expression Omnibus (GEO, NCBI) under the accession number GSE106721.

For details of all other procedures, see *Supplemental Experimental Procedures*.

Author Contributions

J.C.R., B.T., D.G., A.L.B., and D.P.P. performed experiments. T.J.C. provided technical assistance for the calcium flux assay. J.C.R. and M.B. designed experiments, interpreted data, and wrote the manuscript. M.B. directed the study. The authors declare no competing financial interests.

Acknowledgements

J.C.R. was supported by the Canadian Institutes of Health Research (CIHR) Doctoral Award, Michael G. DeGroote Doctoral Scholarship, and an Ontario Graduate Fellowship. D.G. and D.P.P. were supported by the Jans Graduate Scholarship in Stem Cell Research. We acknowledge M. Graham for human sample processing; M.P. Mathieu and A. Fiebig-Comyn for animal husbandry and *in vivo* experimental assistance; Z. Shapovalova, M. Subapanditha, and N. McFarlane for their assistance with cell sorting experiments; and Dr. S. Raha and M. Wong for procedures related to hypoxia. We also thank members of the Labor and Delivery Ward at the McMaster Children's Hospital for supplying human cord blood samples.

References

Balkwill, F. (2004). Cancer and the chemokine network. Nat Rev Cancer 4, 540-550.

Boyd, A.L., and Bhatia, M. (2014). Bone marrow localization and functional properties of human hematopoietic stem cells. Current opinion in hematology 21, 249-255.

Boyd, A.L., Reid, J.C., Salci, K.R., Aslostovar, L., Benoit, Y.D., Shapovalova, Z., Nakanishi, M., Porras, D.P., Almakadi, M., Campbell, C.J.V., et al. (2017). Acute myeloid leukaemia disrupts endogenous myelo-erythropoiesis by compromising the adipocyte bone marrow niche. Nat Cell Biol.

Brenner, S., Whiting-Theobald, N., Kawai, T., Linton, G.F., Rudikoff, A.G., Choi, U., Ryser, M.F., Murphy, P.M., Sechler, J.M., and Malech, H.L. (2004). CXCR4-transgene expression significantly improves marrow engraftment of cultured hematopoietic stem cells. Stem cells (Dayton, Ohio) 22, 1128-1133.

Broxmeyer, H.E., Mejia, J.A., Hangoc, G., Barese, C., Dinauer, M., and Cooper, S. (2007). SDF-1/CXCL12 enhances in vitro replating capacity of murine and human multipotential and macrophage progenitor cells. Stem cells and development 16, 589-596.

Broxmeyer, H.E., Orschell, C.M., Clapp, D.W., Hangoc, G., Cooper, S., Plett, P.A., Liles, W.C., Li, X., Graham-Evans, B., Campbell, T.B., et al. (2005). Rapid mobilization of murine and human hematopoietic stem and progenitor cells with AMD3100, a CXCR4 antagonist. The Journal of experimental medicine 201, 1307-1318.

Chadwick, K., Wang, L., Li, L., Menendez, P., Murdoch, B., Rouleau, A., and Bhatia, M. (2003). Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. Blood 102, 906-915.

Copeland, E. (2006). Hematopoietic Stem-Cell Transplantation. The New England Journal of Medicine 354, 1813-1826.

Crans, H., and Sakamoto, K. (2001). Transcription factors and translocations in lymphoid and myeloid leukemia. Leukemia 15, 313-331.

Cutler, C., Multani, P., Robbins, D., Kim, H.T., Le, T., Hoggatt, J., Pelus, L.M., Desponts, C., Chen, Y.B., Rezner, B., et al. (2013). Prostaglandin-modulated umbilical cord blood hematopoietic stem cell transplantation. Blood 122, 3074-3081.

Delaney, C., Heimfeld, S., Brashem-Stein, C., Voorhies, H., Manger, R.L., and Bernstein, I.D. (2010). Notch-mediated expansion of human cord blood progenitor cells capable of rapid myeloid reconstitution. Nat Med 16, 232-236.

Doulatov S, Vo LT, Macari ER, Wahlster L, Kinney MA, Taylor AM, Barragan J, Gupta M, McGrath K, Lee HY, et al. (2017). Drug discovery for Diamond-Blackfan anemia using reprogrammed hematopoietic progenitors. Sci Transl Med 9.

Doulatov, S., Vo, L., Chou, S., Kim, P., Arora, N., Li, H., Hadland, B., Bernstein, I., Collins, J., Zon, L., et al. (2013). Induction of Multipotential Hematopoietic Progenitors from Human Pluripotent Stem Cells via Respecification of Lineage-Restricted Precursors. Cell stem cell 13, 459-470.

Gluckman, E. (2009). History of cord blood transplantation. Bone Marrow Transplant 44, 621-626.

Goessling, W., Allen, R.S., Guan, X., Jin, P., Uchida, N., Dovey, M., Harris, J.M., Metzger, M.E., Bonifacino, A.C., Stroncek, D., et al. (2011). Prostaglandin E2 Enhances Human Cord Blood Stem Cell Xenotransplants and Shows Long-Term Safety in Preclinical Nonhuman Primate Transplant Models. Cell stem cell 8, 445-458.

Gori, J.L., Butler, J.M., Chan, Y.-Y., Chandrasekaran, D., Poulos, M.G., Ginsberg, M., Nolan, D.J., Elemento, O., Wood, B.L., Adair, J.E., et al. (2015). Vascular niche promotes hematopoietic multipotent progenitor formation from pluripotent stem cells. J Clin Invest 125, 1243-1254.

Gratwohl, A., Pasquini, M., Aljurf, M., Atsuta, Y., Baldomero, H., Foeken, L., Gratwohl, M., Bouzas, L., Confer, D., Frauendorfer, K., et al. (2015). One million haemopoietic stem-cell transplants: a retrospective observational study. Lancet Hematol 2, e91-100.

Guo, B., Huang, X., Cooper, S., and Broxmeyer, H.E. (2017). Glucocorticoid hormone-induced chromatin remodeling enhances human hematopoietic stem cell homing and engraftment. Nat Med.

Heusinkveld, L.E., Yim, E., Yang, A., Azani, A.B., Liu, Q., Gao, J.L., McDermott, D.H., and Murphy, P.M. (2017). Pathogenesis, diagnosis and therapeutic strategies in WHIM syndrome immunodeficiency. Expert Opin Orphan Drugs 5, 813-825.

Hofmeister, C.C., Zhang, J., Knight, K.L., Le, P., and Stiff, P.J. (2007). Ex vivo expansion of umbilical cord blood stem cells for transplantation: growing knowledge from the hematopoietic niche. Bone Marrow Transplant 39, 11-23.

Kahn, J., Byk, T., Jansson-Sjostrand, L., Petit, I., Shivtiel, S., Nagler, A., Hardan, I., Deutsch, V., Gazit, Z., Gazit, D., et al. (2004). Overexpression of CXCR4 on human CD34+ progenitors increases their proliferation, migration, and NOD/SCID repopulation. Blood 103, 2942-2949.

Lapidot, T., and Kollet, O. (2002). The essential roles of the chemokine SDF-1 and its receptor CXCR4 in human stem cell homing and repopulation of transplanted immune-deficient NOD/SCID and NOD/SCID/B2m(null) mice. Leukemia 16, 1992-2003.

Ledran, M.H., Krassowska, A., Armstrong, L., Dimmick, I., Renström, J., Lang, R., Yung, S., Santibanez-Coref, M., Dzierzak, E., Stojkovic, M., et al. (2008). Efficient Hematopoietic Differentiation of Human Embryonic Stem Cells on Stromal Cells Derived from Hematopoietic Niches. Cell stem cell 3, 85-98.

Lee, J., Dykstra, B., Spencer, J.A., Kenney, L.L., Greiner, D.L., Shultz, L.D., Brehm, M.A., Lin, C.P., Sackstein, R., and Rossi, D.J. (2017). mRNA-mediated glycoengineering ameliorates deficient homing of human stem cell-derived hematopoietic progenitors. J Clin Invest.

Nagasawa, T., Hirota, S., Tachibana, K., Takakura, N., Nishikawa, S., Kitamura, Y., Yoshida, N., Kikutani, H., and Kishimoto, T. (1996). Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. Nature 382, 635-638.

Nakano, H., Lyons-Cohen, M.R., Whitehead, G.S., Nakano, K., and Cook, D.N. (2017). Distinct functions of CXCR4, CCR2, and CX3CR1 direct dendritic cell precursors from the bone marrow to the lung. J Leukoc Biol 101, 1143-1153.

Ng, E.S., Azzola, L., Bruveris, F.F., Calvanese, V., Phipson, B., Vlahos, K., Hirst, C., Jokubaitis, V.J., Yu, Q.C., Maksimovic, J., et al. (2016). Differentiation of human embryonic stem cells to HOXA+ hemogenic vasculature that resembles the aorta-gonad-mesonephros. Nature biotechnology 34, 1168-1179.

Ramos-Mejia, V., Navarro-Montero, O., Ayllon, V., Bueno, C., Romero, T., Real, P.J., and Menendez, P. (2014). HOXA9 promotes hematopoietic commitment of human embryonic stem cells. Blood 124, 3065-3075.

Ran, D., Shia, W.J., Lo, M.C., Fan, J.B., Knorr, D.A., Ferrell, P.I., Ye, Z., Yan, M., Cheng, L., Kaufman, D.S., et al. (2013). RUNX1a enhances hematopoietic lineage commitment from human embryonic stem cells and inducible pluripotent stem cells. Blood 121, 2882-2890.

Risueño, R.M., Sachlos, E., Lee, J.-H., Lee, J.B., Hong, S.-H., Szabo, E., and Bhatia, M. (2012). Inability of Human Induced Pluripotent Stem Cell-Hematopoietic Derivatives to Downregulate MicroRNAs In Vivo Reveals a Block in Xenograft Hematopoietic Regeneration. Stem cells (Dayton, Ohio) 30, 131-139.

Rodriguez-Lara, V., Ignacio, G.S., and Cerbon Cervantes, M.A. (2017). Estrogen induces CXCR4 overexpression and CXCR4/CXL12 pathway activation in lung adenocarcinoma cells in vitro. Endocr Res, 1-13.

Salvagiotto, G., Zhao, Y., Vodyanik, M., Ruotti, V., Stewart, R., Marra, M., Thomson, J., Eaves, C., and Slukvin, I. (2008). Molecular profiling reveals similarities and differences between primitive subsets of hematopoietic cells generated in vitro from human embryonic stem cells and in vivo during embryogenesis. Experimental hematology 36, 1377-1389.

Saxena, S., Ronn, R.E., Guibentif, C., Moraghebi, R., and Woods, N.B. (2016). Cyclic AMP Signaling through Epac Axis Modulates Human Hemogenic Endothelium and Enhances Hematopoietic Cell Generation. Stem Cell Reports.

Scheurer, S.B., Rybak, J.N., Rosli, C., Neri, D., and Elia, G. (2004). Modulation of gene expression by hypoxia in human umbilical cord vein endothelial cells: A transcriptomic and proteomic study. Proteomics 4, 1737-1760.

Sugimura, R., Jha, D.K., Han, A., Soria-Valles, C., da Rocha, E.L., Lu, Y.-F., Goettel, J.A., Serrao, E., Rowe, R.G., Malleshaiah, M., et al. (2017). Haematopoietic stem and progenitor cells from human pluripotent stem cells. Nature.

Sugiyama, T., Kohara, H., Noda, M., and Nagasawa, T. (2006). Maintenance of the Hematopoietic Stem Cell Pool by CXCL12-CXCR4 Chemokine Signaling in Bone Marrow Stromal Cell Niches. Immunity 25, 977-988.

Suzuki, N., Yamazaki, S., Yamaguchi, T., Okabe, M., Masaki, H., Takaki, S., Otsu, M., and Nakauchi, H. (2013). Generation of Engraftable Hematopoietic Stem Cells From Induced Pluripotent Stem Cells by Way of Teratoma Formation. Molecular Therapy 21, 1424-1431.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43, D447-452.

Tian, X., Woll, P.S., Morris, J.K., Linehan, J.L., and Kaufman, D.S. (2006). Hematopoietic engraftment of human embryonic stem cell-derived cells is regulated by recipient innate immunity. Stem cells (Dayton, Ohio) 24, 1370-1380.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science (New York, NY) 347, 1260419.

Vodyanik, M., Thomson, J.A., and Slukvin, I. (2006). Leukosialin (CD43) defines hematopoietic progenitors in human embryonic stem cell differentiation cultures. Blood 108, 2095-2105.

Wang, L.M., P.; Shojaei, F.; Li, L.; Mazurier, F.; Dick, J. E.; Cerdan, C.; Levac, K.; Bhatia, M. (2005). Generation of hematopoietic repopulating cells from human embryonic stem cells independent of ectopic HOXB4 expression. Journal of Experimental Medicine 201, 1603-1614.

Wang, W., Yu, S., Myers, J., Wang, Y., Xin, W.W., Alkabri, M., Xin, A.W., Li, M., Huang, A.Y., Xin, W., et al. (2017). Notch2 blockade enhances hematopoietic stem cell mobilization and homing. Haematologica.

Weisel, K.C., Gao, Y., Shieh, J.-H., and Moore, M.A.S. (2006). Stromal cell lines from the aorta-gonado-mesonephros region are potent supporters of murine and human hematopoiesis. Experimental hematology 34, 1505-1516.

Wright, D.E., Bowman, E.P., Wagers, A.J., Butcher, E.C., and Weissman, I.L. (2002). Hematopoietic stem cells are uniquely selective in their migratory response to chemokines. Journal of Experimental Medicine 195, 1145-1154.

Zhang, W.B., Navenot, J.M., Haribabu, B., Tamamura, H., Hiramatu, K., Omagari, A., Pei, G., Manfredi, J.P., Fujii, N., Broach, J.R., et al. (2002). A point mutation that confers constitutive activity to CXCR4 reveals that T140 is an inverse agonist and that AMD3100 and ALX40-4C are weak partial agonists. J Biol Chem 277, 24515-24521.

Figures and Tables

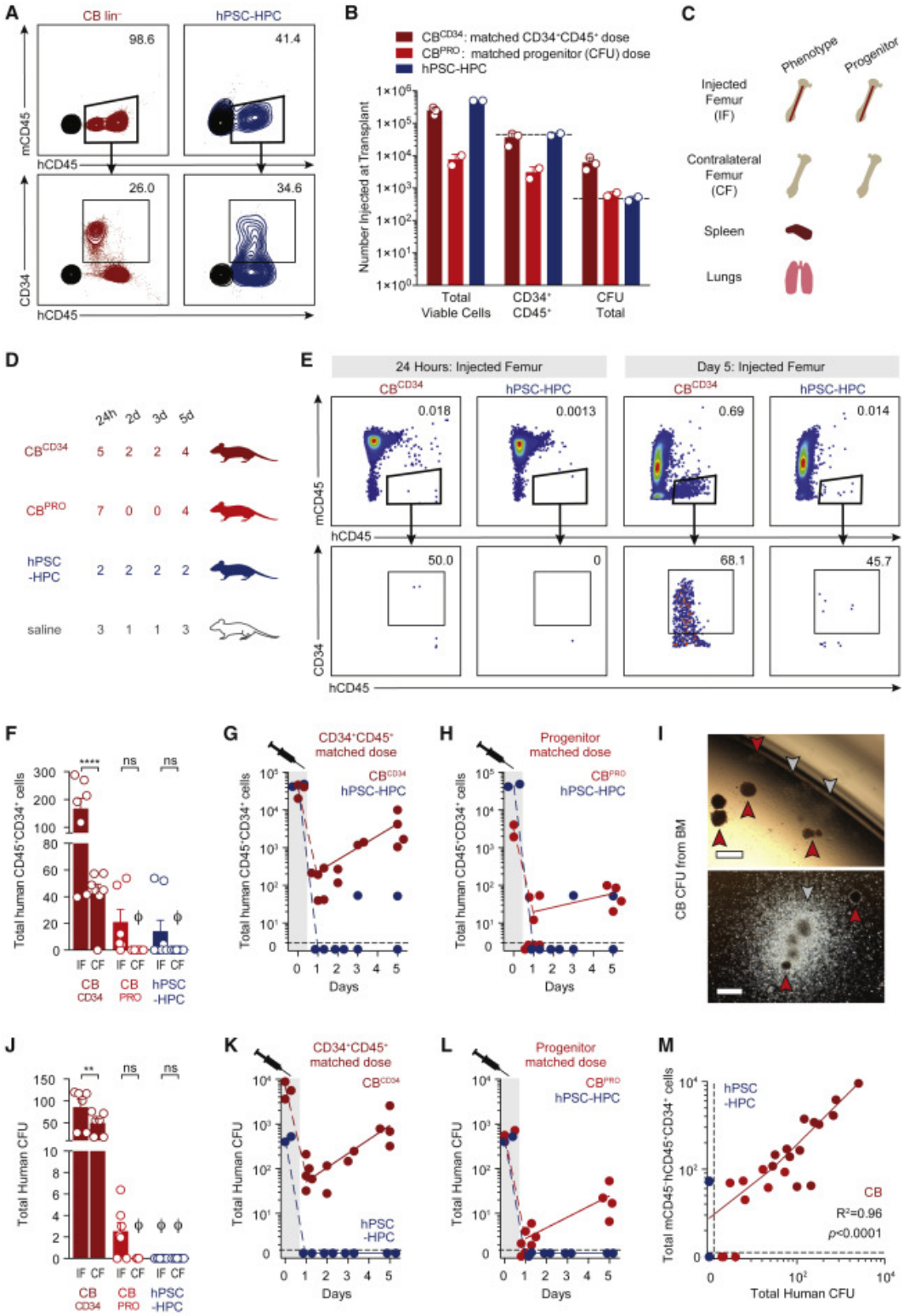


Figure legend on next page

Figure 1. Transplantation kinetics of somatic and hPSC-derived HPCs

(A) Phenotype of CB and hPSC-derived HPCs in injection fraction (human $CD34^+$ human $CD45^+$ mouse $CD45^-$), analysed on the day of transplant. Mouse CD45 antibody use validated to exclude mouse cells upon BM transplantation. (B) Total viable cells, and total phenotypic ($CD34^+CD45^+$) and functional (CFU) CB and hPSC-derived HPCs in injection fraction, analysed on the day of transplant. CB^{CD34} and CB^{PRO} cell doses control for input phenotypic and functional hPSC-HPCs, respectively. *N*=2-3 independent transplants, each dot is an independent transplant. (C) Tissue allocation for phenotypic (flow cytometry) and functional (CFU) assays. (D) Transplant groups, *N*=3 total independent transplants, *n*=6 harvest analyses. (E) Phenotype of CB and hPSC-derived HPCs from harvested BM. (F) Total $mCD45^-$ $hCD45^+CD34^+$ cells retained in the BM of injected (IF) and contralateral (CF) femurs. To assess BM retention separately from cellular proliferation and expansion, only 24 and 48 h retention data for CB shown; day 3 and 5 data omitted. Each data point is one mouse, Ø is zero for group. Two-way ANOVA, p<0.0001****. (G-H) Total $mCD45^-hCD45^+CD34^+$ cells per injected femur. Same hPSC-HPC data in both panels. (I) CFU from CB-transplanted BM, harvested at day 5. Arrowheads, red = burst-forming unit-erythroid; grey = colony forming unit CFU-granulocyte, -monocyte, or -granulocyte/monocyte. (J) Total human CFU per harvested IF and CF BM. To assess BM retention of progenitors separately from cellular proliferation and expansion, only 24 and 48 h retention data for CB shown; day 3 and 5 data omitted. Each data point is one mouse, Ø is zero for group. One-way ANOVA, p<0.01**. (K-L) Total human CFU per IF. Same hPSC-HPC data in both panels. (M) Linear regression of total CB phenotypic versus functional HPCs quantified per IF (*n*=24 mice). hPSC-HPC data also shown (*n*=8 mice). Each data point is one mouse.

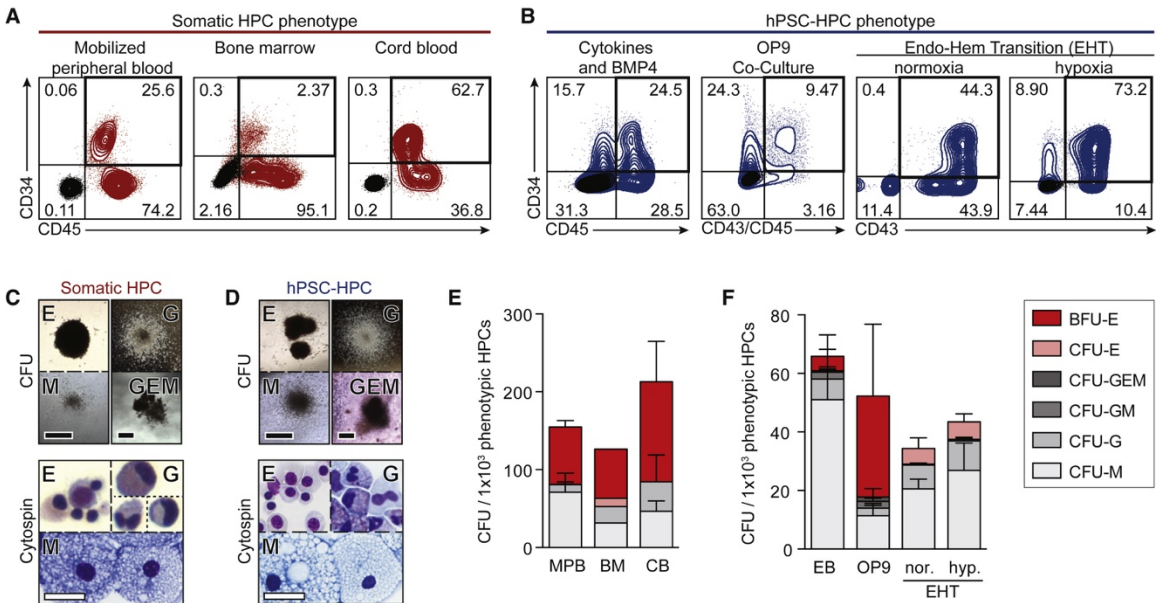


Figure 2. Shared progenitor capacity of somatic and hPSC-derived HPCs

(A) Phenotypic ($CD34^+CD45^+$) somatic HPCs, by flow cytometry. Negative stain in black. (B) Phenotypic ($CD34^+$ with $CD45^+$ and/or $CD43^+$) hPSC-HPCs, by flow cytometry, with: Cytokines and BMP4 (Chadwick et al., 2003); OP9 Co-Culture (Vodyanik et al., 2006); and EHT in normoxia or hypoxia (Lee et al., 2017a) differentiation methods. Negative stain in black. (C-D) Representative somatic (C), and hPSC-derived (D) hematopoietic colonies (CFU); picked CFU were spun on to slides for Wright-Giemsa staining (Cytospin). White scale bar is 500 µm, black scale bar is 30 µm. (E-F) Total CFU relative to $1.0x10^3$ input phenotypic HPCs from (E) somatic HPCs: MPB, *N*=3 (healthy donors), *n*=8 (biological replicates); BM, *N*=1, *n*=4; and CB, *N*=2, *n*=4, and (F) hPSC-HPCs: Cytokines and BMP4, *N*=4 (independent experiment), *n*=10 (biological replicates); OP9 Co-Culture, *N*=4, *n*=10; EHT supernatant (sup.) cells, *N*=2, *n*=6; and EHT adherent (adh.) cells, *N*=2, *n*=6 cultured in normoxia.

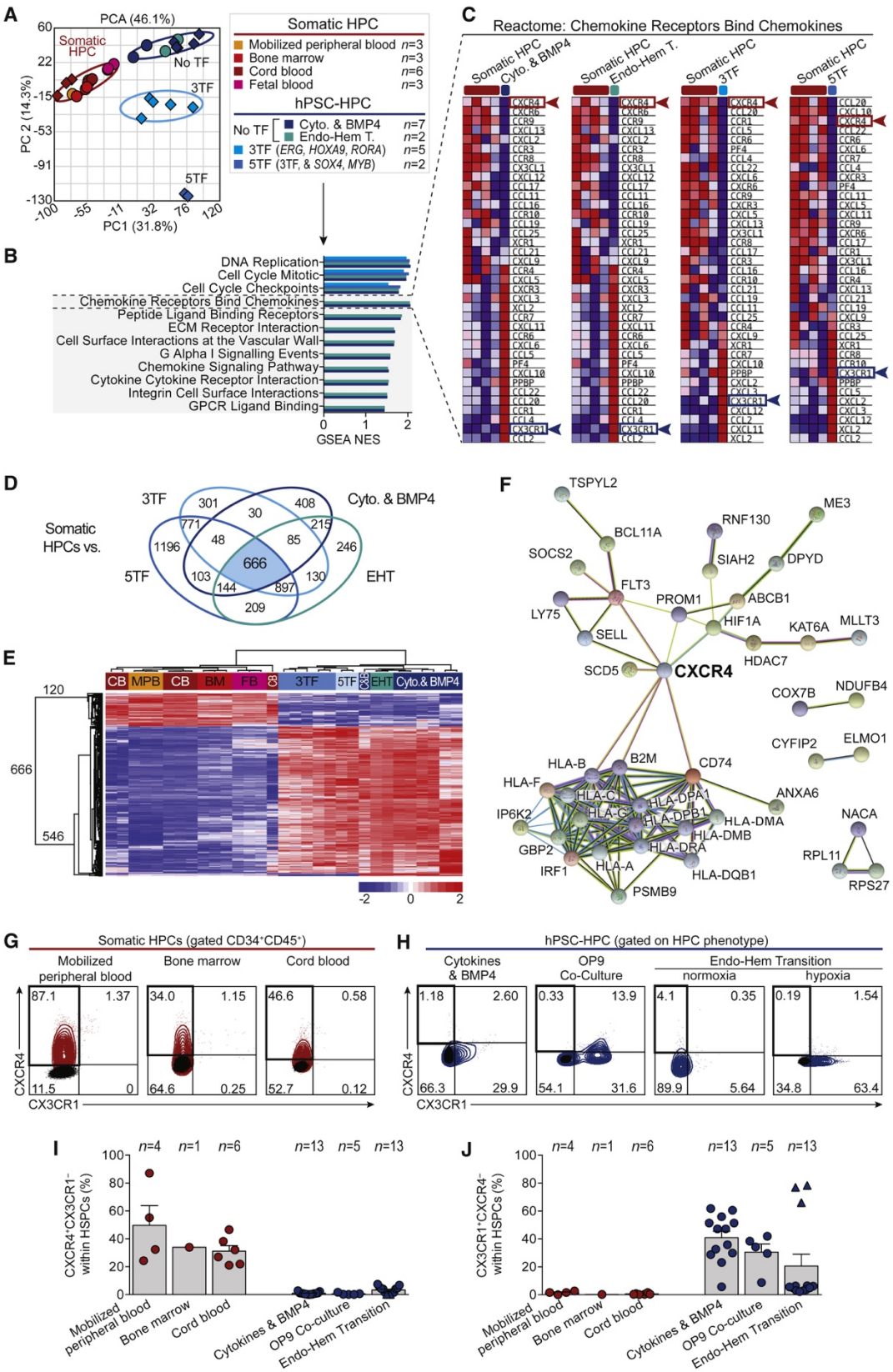


Figure legend on next page

Figure 3. Identification of aberrant chemokine receptor expression by hPSC-HPCs

(A) Principal component analysis (PCA) of global transcriptome from FACS-purified $CD45^{+}CD34^{+}CD38^{\pm}$ HPCs, including samples from GSE49938 (Doulatov et al., 2013), and from our lab (GSE3823 and GSE106721), diamond and circle symbols, respectively. Further details in Supplemental Experimental Procedures. (B) Gene sets identified by GSEA as enriched in hPSC- versus somatic HPCs (*N*=15), using meaned groups described in Figure 1A. GSEA reports in Table S1. (C) Blue-Pink O' Gram heat maps from GSEA report using meaned groups described in Figure 1A. (D) ANOVA comparing the four sources of hPSC-HPCs individually to somatic HPCs (MPB, BM, CB, and FB), depicted as a Venn diagram. (E) Unsupervised hierarchical clustering of hPSC- and somatic HPCs using 666 genes identified in Figure 3D. (F) STRING analysis of 120 somatic HPC genes, where 43 are shown. 77 disconnected nodes were removed. (G-H) Representative flow cytometry of CXCR4 and CX3CR1 of $CD34^{+}CD45^{+}$ somatic HPCs (G), and hPSC-HPCs (H; HPC phenotypes in Figure 2B). Negative stain in black. (I-J) Summary of $CXCR4^{+}CX3CR1^{-}$ (I), and $CX3CR1^{+}CXCR4^{-}$ (J), expression of phenotypic HPCs, as assessed by flow cytometry. Each dot is an independent healthy donor or biological replicate from somatic and hPSC-derived HPCs, respectively. Triangle symbols indicate biological replicates cultured in hypoxia.

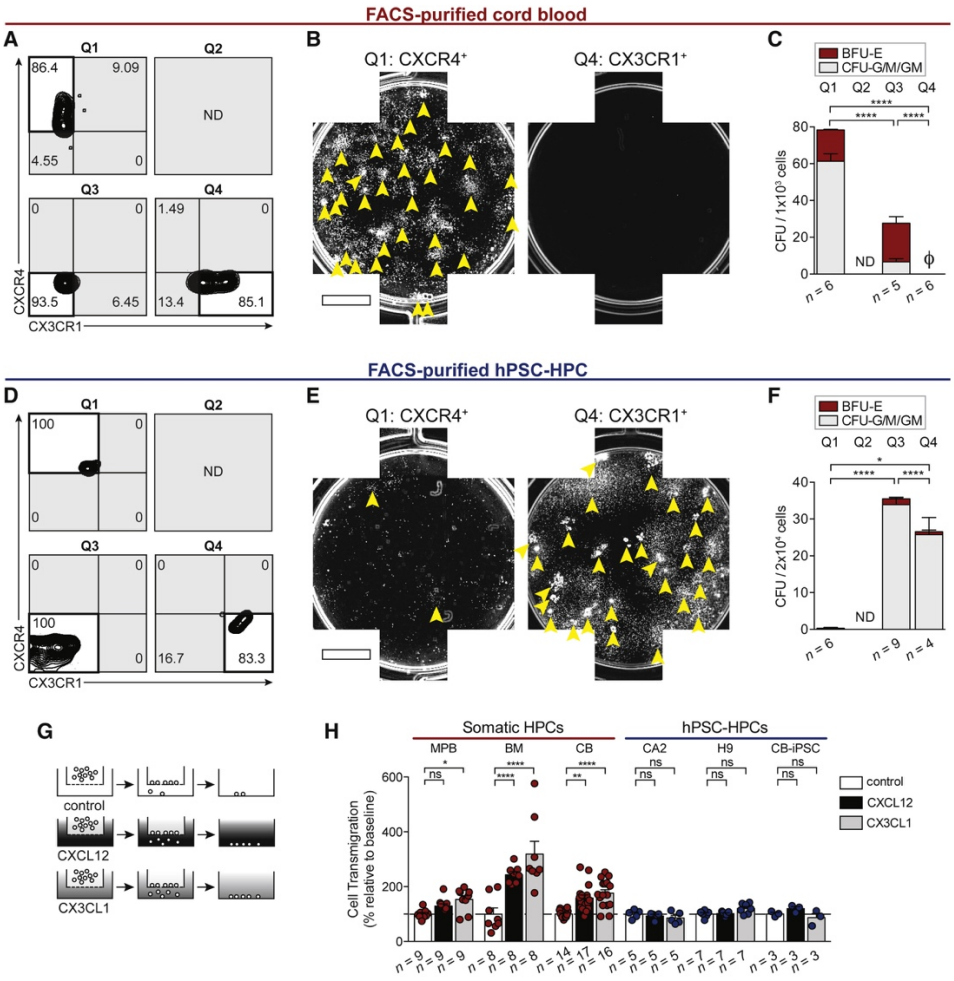


Figure 4. Differential segregation of progenitors by CXCR4 and CX3CR1

(A) Sort purities from CB cells FACS-purified by CXCR4 and CX3CR1 expression. Quadrant (Q) 2 was not detected (ND). (B) Composite well images of Q1- and Q4-sorted CB. $5x10^2$ Q1 and $1x10^3$ Q4 sorted cells were seeded per well. Scale bar represents 5 mm. (C) CFU frequency of FACS-purified CB based on chemokine receptor expression. *N*=1 healthy donor, *n*=5-6 (biological replicates per quadrant). Ø is zero for group. Q2 was ND. (D) Sort purities from hPSC-derived hematopoietic cells FACS-purified by CXCR4 and CX3CR1 expression. Q2 was ND. (E) Composite well images of Q1- and Q4-sorted hPSC-derived hematopoietic cells. $2x10^4$ sorted cells were seeded per well. Scale bar represents 5 mm. (F) CFU frequency of FACS-purified hPSC-derived hematopoietic cells based on chemokine receptor expression. *N*=2 (independent experiments), *n*=4-9. Q2 was ND. (G) Transwell assay was conducted with 200 ng/mL CXCL12, or 200 ng/mL CX3CL1, or control (0.001% BSA). (H) Transwell migration was quantified by flow cytometry. MPB *N*=3 healthy donors, *n*=9-10 biological replicates; BM *N*=1, *n*=8; CB *N*=5, *n*=14-17; and hPSC-HPCs *N*=5: CA2 *N*=2 independent experiments, *n*=5-6; H9 *N*=2, *n*=6; and CB-iPSC *N*=1, *n*=3. CB used as positive control in every hPSC-HPC experiment. Two-way ANOVA, $p<0.05$*, $P<0.01$**, $p<0.0001$****.

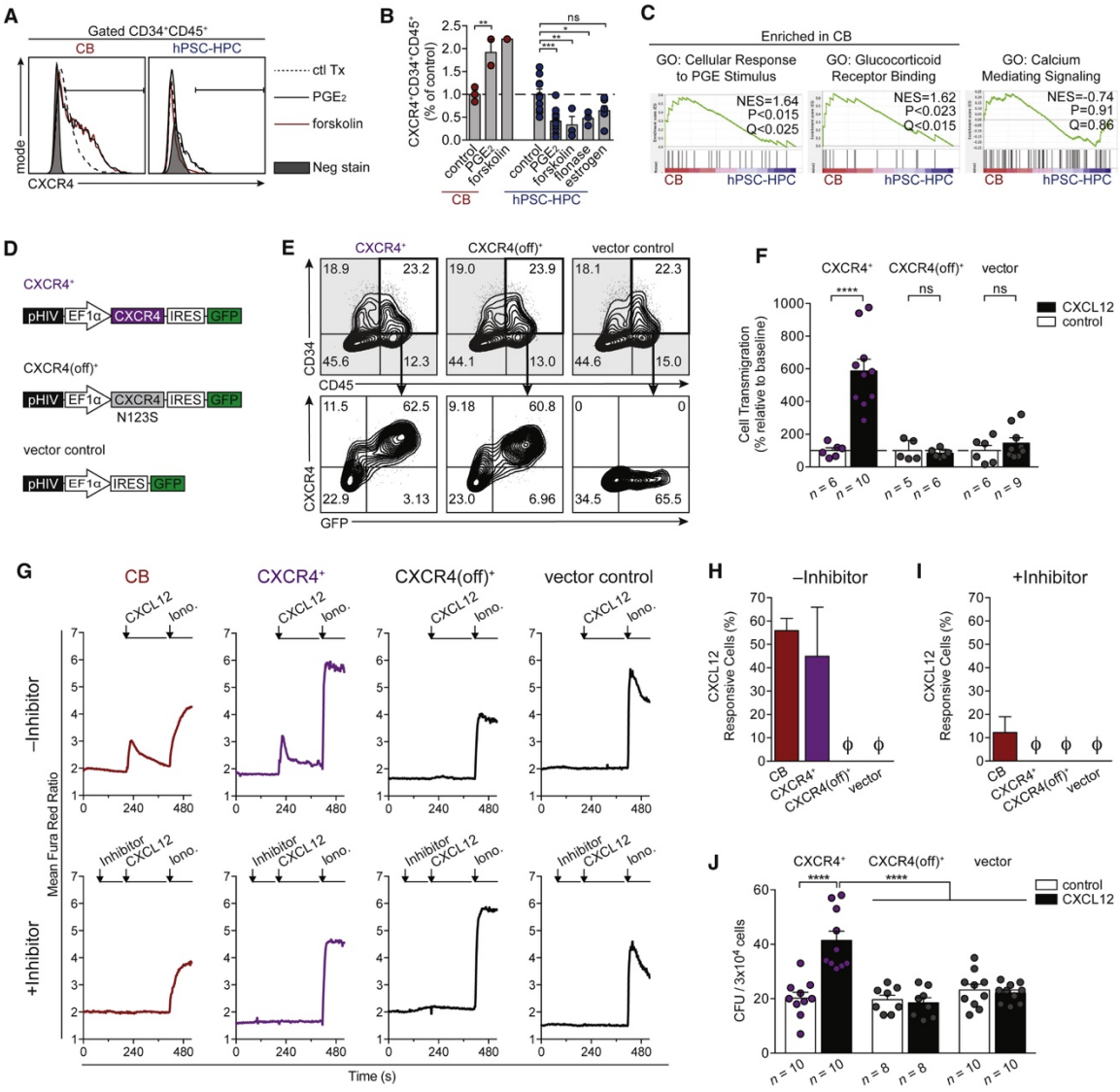


Figure 5. Overcoming resistance to CXCR4 expression enables CXCL12-dependent signaling

(A) Flow cytometric of CXCR4 within CD34⁺CD45⁺ CB or hPSC-HPCs treated with indicated compounds; see also Supplemental Experimental Procedures. (B) A summary of total CXCR4⁺CD34⁺CD45⁺ cells from CB (*N*=2 experiments) and hPSC-HPCs (*N*=3), relative to control (0.01% DMSO or BSA). Each dot is a biological replicate. (C) GSEA analysis of pathways which regulate CXCR4. Samples: CB, *N*=3, H9-HPC, *N*=2; CD34⁺CD45⁺CD38⁻, shown in Figure 1A. *Abbreviations:* Gene Ontology, GO; normalized enrichment score, NES. (D) Human *CXCR4* was cloned into pHIV-EGFP, which has an internal ribosome entry site (IRES) translation link to GFP (CXCR4⁺). Site-directed mutagenesis of *CXCR4* (N123K) was termed CXCR4(off)⁺. Vector control, expressing GFP, was used in parallel for all experiments. (E) Flow cytometry at 48 h post-transduction on EB day 16, showing comparable hPSC-HPC frequency (CD34⁺CD45⁺), and robust CXCR4⁺GFP⁺ co-expression. (F) Transwell assay was conducted with 200 ng/mL CXCL12 or control (0.001% BSA), and quantified by flow

cytometry at 48 h post-transduction on EB day 16. CXCR4$^+$ *N*=3 independent experiments, *n*=6-10 biological replicates; CXCR4(off)$^+$ *N*=2, *n*=5-6; vector control, *N*=3, *n*=6-9. Two-way ANOVA, p<0.0001****. (G) Calcium flux transients were monitored in response to CXCL12 (200 ng/mL) in the absence (*top row*) or presence (*bottom row*) of CXCR4 inhibitor (AMD3100, 10 μM) at 48 h post-transduction on EB day 16. Ionomycin (10uM) treatment was used as a positive control to identify live cells. (H-I) The frequency of cells that responded to CXCL12 treatment (transient > 150% of baseline) was quantified in the absence (H) or presence (I) of CXCR4 inhibitor, AMD3100. (H) CB: *N*=3 healthy donors, *n*=5 biological replicates; all 3 hPSC-HPC groups: *N*=2 independent experiments, *n*=3-4. Ø is zero for group. (I) CB, *N*=3, *n*=3; all 3 hPSC-HPC groups, *N*=2, *n*=4. (J) Total CFU counts from transduced hPSC-HPCs seeded into MethoCult ±150 ng/mL CXCL12 at 48 h post-transduction on EB day 16. CXCR4$^+$ and vector control: *N*=3, *n*=10; and CXCR4(off)$^+$: *N*=2, *n*=8. Two-way ANOVA, p<0.01**, p<0.0001****.

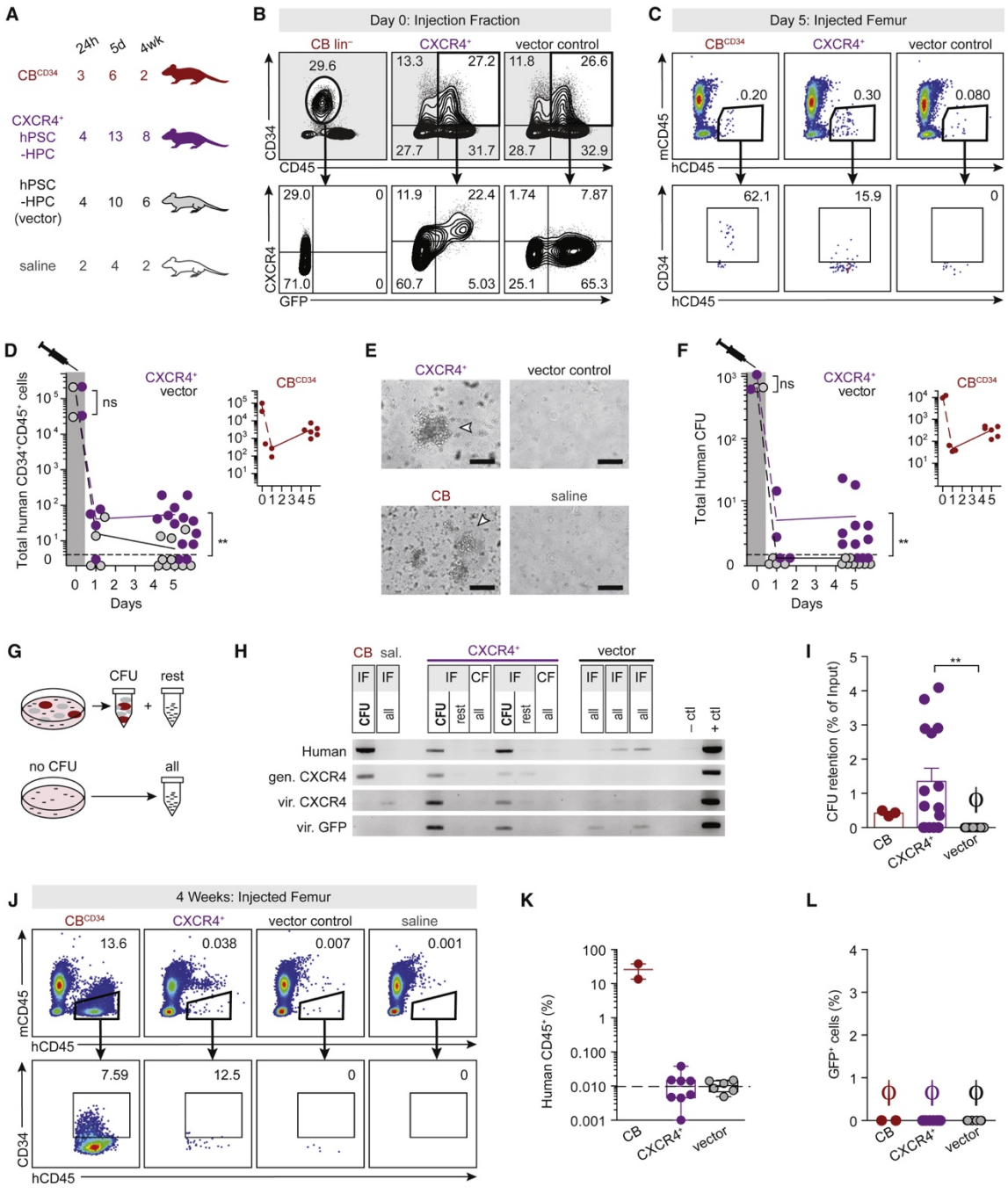


Figure 6. CXCR4 rescues transplantation deficiency of hPSC-HPCs

(A) Transplant groups: 5x10⁴ CD34⁺CD45⁺ cells were transplanted from CB^{CD34} and both hPSC-HPC groups (48 h post-transduction, on EB day 16). BM from saline-injected mice were collected at every harvest. *N*=2 independent transplants, *n*=3 harvest analyses. (B) Injection fraction phenotyping. (C) Flow cytometry of BM harvested on day 5. (D) Total mCD45⁻ hCD45⁺CD34⁺ cells per IF. CXCR4⁺ and vector in purple and black, respectively. CB in inset (red). Each data point is one mouse. Two-tailed T Test, p<0.01**. (E) CFU analysis from harvested BM, scale bar is 100 μm. (F) Total human CFU per IF. CXCR4⁺ and vector in purple

and black, respectively. CB in inset (red). Each data point represents one mouse. Two-tailed T Test, $p<0.01$**. (G) 1-5 colonies (2±2) per transplanted BM samples were picked by micropipette, and the non-colony remainder of the well was kept separately ('rest'). If no CFU were observed, the entire well was collected ('all'). (H) CFU were analyzed for human and vector sequences by PCR. Samples from BM harvested on day 5. Each box is one NSG mouse. PCR was completed with 1-5 colonies (2±2) per sample. Two representative CXCR4+ mice are shown, with 2 (left) and 1 (right) colony shown. CB PCR was completed with 1 colony. Genomic, gen.; viral, vir.; and saline, sal. (I) CFU retention per IF as a percentage of input CFU. Each data point is one mouse. To assess BM retention separately from cellular proliferation and expansion, only 24 h retention data for CB shown; day 5 data omitted. Ø is zero for group. (J) Flow cytometry of CB and hPSC-HPC transplanted IF, harvested 4 wk post-transplant. (K) Human chimerism ($hCD45^{+}mCD45^{-}$) at 4 wk. (L) GFP frequency at 4 wk.

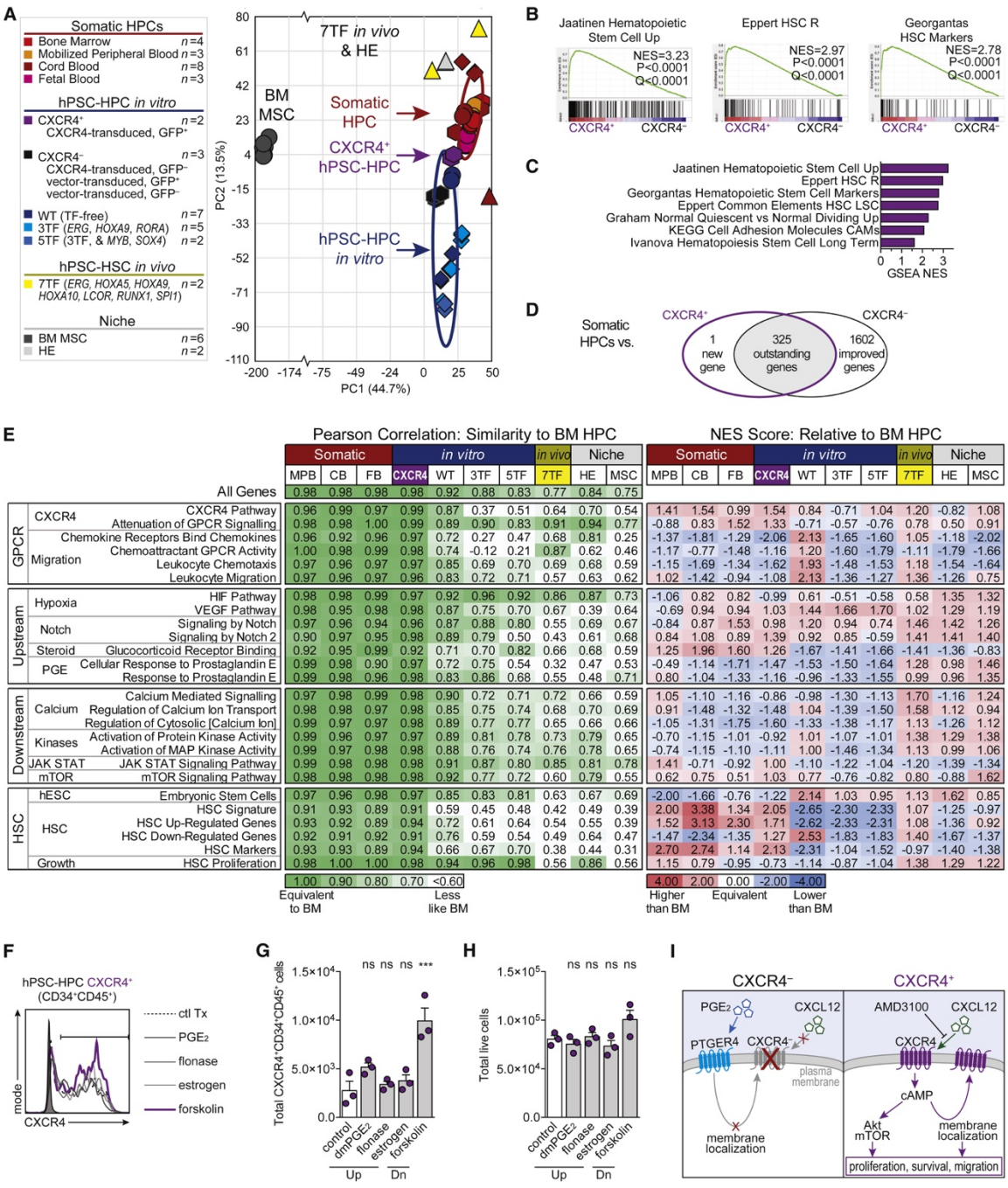


Figure 7. CXCR4 auto-regulation as a target for hPSC-HSC development

(A) PCA correlation biplot comparing the global transcriptome of FACS-purified HPCs, CD45⁺CD34⁺CD38± from a combined dataset including GSE83719 (triangles), GSE49938 (diamonds), from our lab (GSE3823 and GSE92778; circles), and original data from this study (hexagon symbols, GSE106721): CD34⁺CD45⁺CXCR4⁺ *N*=2 and CD34⁺CD45⁺CXCR4⁻ *N*=3. Ellipses are 2 standard deviations. (B-C) Enrichment plots (B) and NES scores (C) of gene sets identified by GSEA as enriched in CD34⁺CD45⁺CXCR4⁺ hPSC-HPCs (*N*=2) versus CD34⁺CD45⁺CXCR4⁻ hPSC-HPCs (*N*=3). 195 of 3685 C2 (MSigDB) gene sets were significantly

enriched in $CD34^+CD45^+CXCR4^+$ hPSC-HPCs at nominal p-value < 0.05. (D) ANOVA comparing CXCR4+ and CXCR4- hPSC-HPCs to somatic HPCs (MPB, BM, CB, and FB) is depicted as a Venn diagram. (E) Sample groups described in Figure 7A were assessed at the global level (*All Genes*) using Pearson Correlation Coefficient, and within gene sets, see also Supplemental Experimental Procedures. NES score from GSEA of indicated groups (biological replicates, not averaged) compared to BM HPCs (*N*=3 healthy donors). (F-H) $CXCR4^+$ hPSC-HPCs were exposed to CXCR4-inducing compounds for 48 h; see also Supplemental Experimental Procedures. Representative CXCR4 staining of $CD34^+CD45^+$ cells (F), and quantification of total $CXCR4^+CD34^+CD45^+$ (G), and bulk cells (H). *N*=3 biological replicates. (I) A model of CXCR4 function in hPSC-HPCs.

Supplemental Information

Supplemental figures and tables

Figure S1, related to Figure 1.

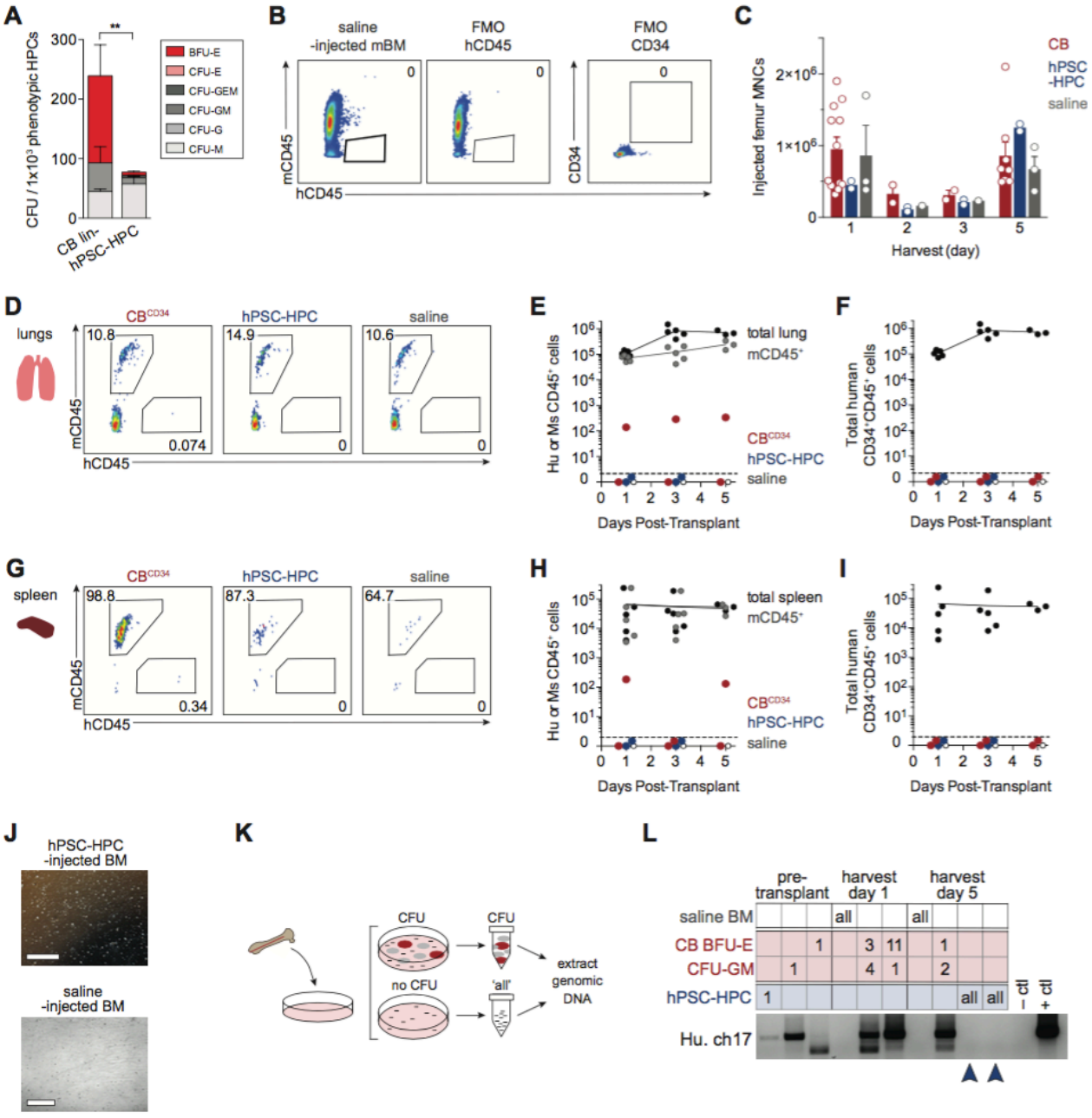


Figure legend on next page

Figure S1. Expanded transplantation kinetics

(A) Total CFU per 1000 phenotypic ($CD34^+CD45^+$) progenitors. *N*=2 independent transplant samples each, tested on Day 0. Burst-forming unit-erythroid (BFU-E), colony forming unit (CFU)-erythroid (E), -granulocyte (G), -monocyte (M), -granulocyte/monocyte (GM), or -mixed lineage (GEM). Unpaired T Test, $p<0.01$**. (B) Flow cytometry of saline-injected BM harvested on day 5, and fluorescence minus one (FMO) controls for hCD45 and CD34 using pooled BM from CB-injected femurs harvested on day 5. Paired CB and hPSC-HPC flow plots shown in Figure 1E. (C) Total BM mononuclear cell (MNC) counts, each dot is one NSG mouse. *N*=3 independent transplants. Two-way ANOVA, no statistical differences. (D) Flow cytometry of lung tissue harvested on day 3. (E-F) Total human and mouse $CD45^+$ cells (E), and human HPCs ($CD34^+hCD45^+mCD45^-$cells; F), per lung sample, *N*=1 experiment. Total lung MNCs in black. Each data point is one mouse. (G) Flow cytometry of mouse spleen tissue harvested on day 3. (H-I) Total human and mouse $CD45^+$ cells (H), and $CD34^+hCD45^+mCD45^-$ cells (I) per spleen, *N*=1 experiment. Total spleen MNCs in black. Each data point is one mouse. (J) No CFU were observed from hPSC-HPC or saline-injected BM, at any harvest. (K) BM MNCs were seeded into MethoCult for CFU analysis. CFU were picked by micropipette. If no CFU were observed, the entire well was collected. Genomic DNA from each sample was extracted separately. (L) CFU were analyzed for human sequences by conventional PCR.

Figure S2, related to Figure 3

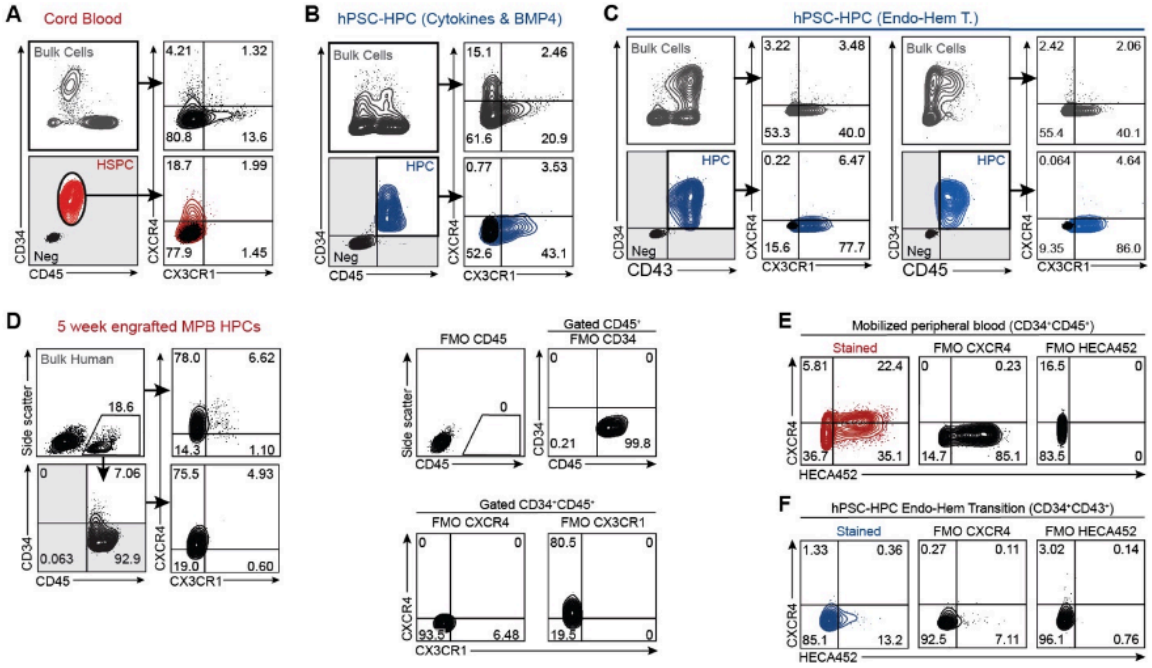


Figure S2. Additional supporting immunophenotyping of somatic and hPSC-derived HPCs

(A-B) Flow cytometry of CXCR4 and CX3CR1 on bulk hematopoietic MNCs (grey) and of CB HPCs (red; A), Cytokines & BMP4-derived HPCs (blue; B), or EHT in hypoxia-derived HPCs (blue; C). Negative stain in black. (D) Immunophenotyping of BM aspirate of NSG mouse transplanted with MPB HPCs for 5 weeks. (E,F) CXCR4 and HECA452 staining of MPB (A; $CD34^+CD45^+$). and EHT-generated hPSC-HPCs (B; $CD34^+CD43^+$). Gates were set with FMO staining using pooled replicate samples.

Figure S3, related to Figure 5

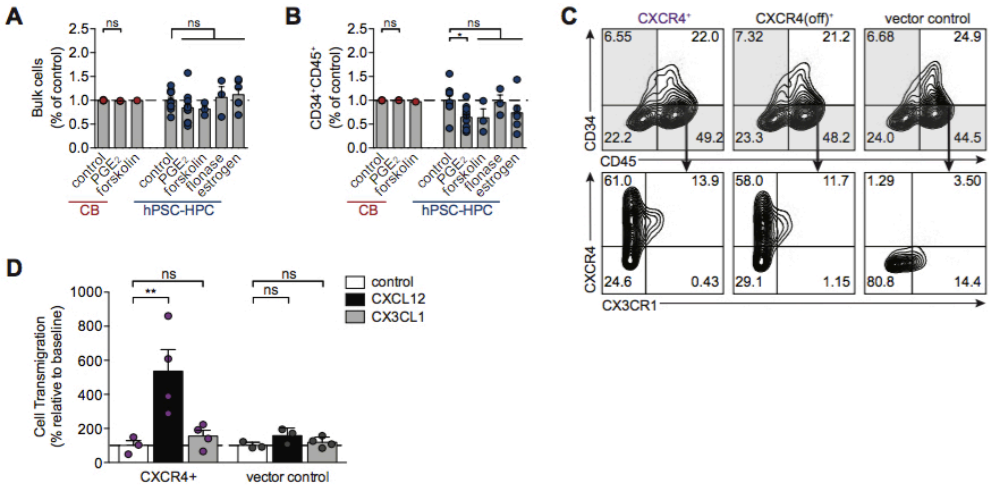


Figure S3. Pharmacological effects on hematopoietic cells

(A-B) CB were treated ±10 μM PGE_2 in 0.01% BSA, or ±10 μM forskolin in 0.1% DMSO in IMDM for 2 h at 37°C. Cells were analysed by flow cytometry for viability (A), and $CD34^+CD45^+$ frequency (B). Differentiating hPSC-derived hematopoietic cells were treated with compounds in to approximate previous methods shown to induce CXCR4 (Table S2), counted, and quantified by flow cytometry for total live cells (A), and HPCs (B; $CD34^+CD45^+$), relative to controls (0.1% DMSO or 0.01% BSA). One-way ANOVA, $p<0.05$*. $n\geq3$. (C) Flow cytometry at 48 h post-transduction on EB day 16, showing comparable hPSC-HPC frequency ($CD34^+CD45^+$), robust $CXCR4^+$ expression, and continued CX3CR1 expression. (D) Transwell assay was conducted with 200 ng/mL CXCL12, 200 ng/mL CX3CL1, or control (0.001% BSA), and quantified by flow cytometry at 48 h post-transduction on EB day 16. n=3-4 biological replicates. Two-way ANOVA, $p<0.01$**. Control and CXCL12 data also shown in Figure 5F.

Figure S4, related to Figure 6

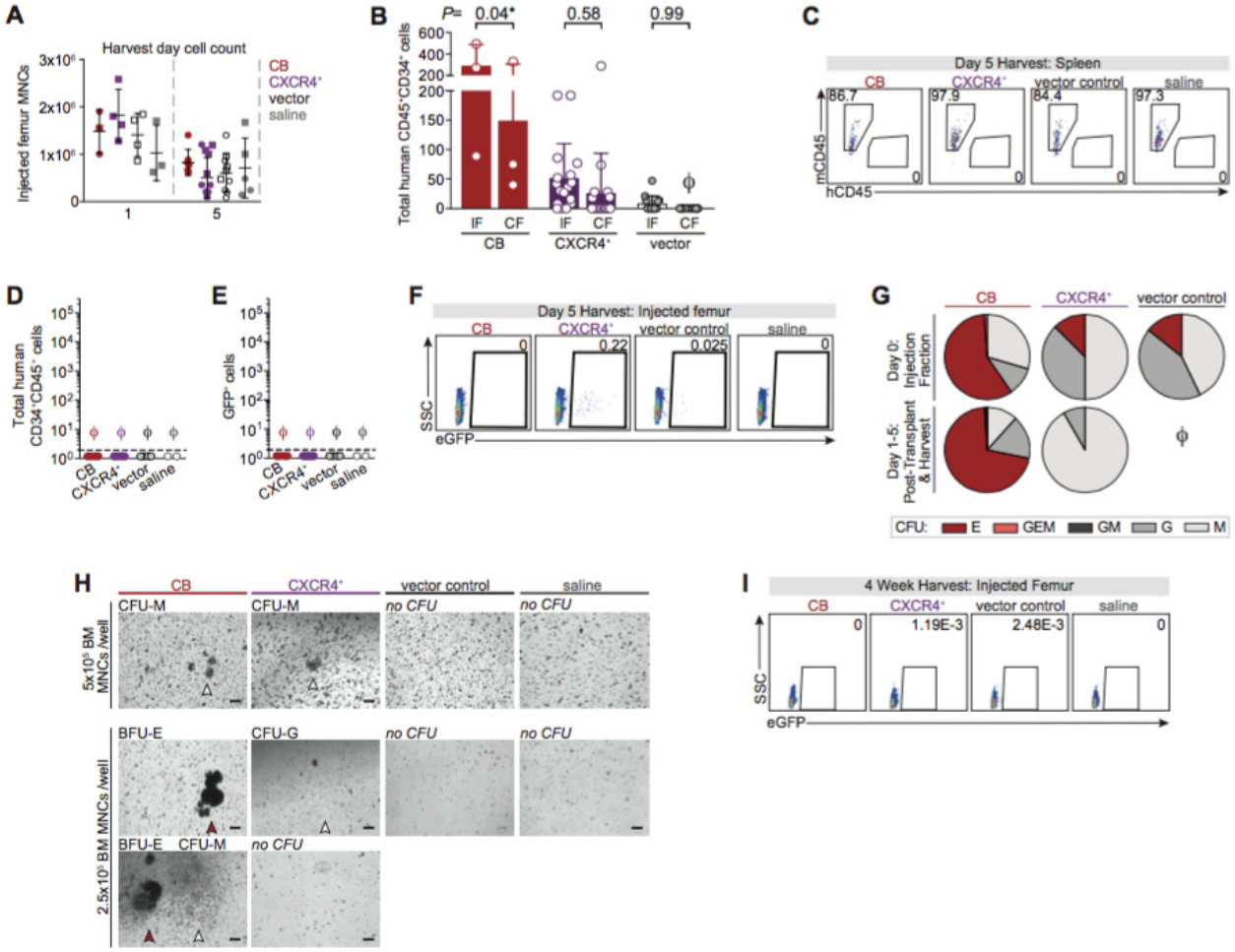


Figure S4. Extended CXCR4⁺ transplant characterization

(A) Total BM counts; each dot is one mouse, *N*=2 independent transplants. Two-way ANOVA. (B) Total mCD45⁻hCD45⁺CD34⁺ cells retained in the BM of injected (IF) and contralateral (CF) femurs. To assess BM retention separately from cellular proliferation and expansion, only 24 h retention data for CB shown; day 5 data omitted. Each data point is one mouse, Ø is zero for group. Two-way ANOVA. (C) Flow cytometry of mouse spleen MNCs harvested on day 5. (D-E) Total CD34⁺hCD45⁺mCD45⁻ cells (D), and GFP⁺ cells (E) per spleen sample. Each data point is one mouse, *N*=1 experiment; CB *n*=4 mice, CXCR4⁺ *n*=10, vector control *n*=4, and saline, *n*=2. Ø is zero for group. (F) GFP analysis of mouse BM MNCs harvested on day 5. (G) CFU lineage at indicated time points. Ø is zero CFU in all vector control mice. (H) Morphology of CFU and background BM debris. CFU-G/M and BFU-E indicated by white and red arrowheads, respectively. Initial cell dose per 6-well is indicated on the left. (I) GFP analysis of mouse BM MNCs harvested at 4 wk.

Table S1. GSEA reports comparing hPSC-HPCs to somatic HPCs

Table S2. ANOVA Results of 666 Consistently and Differentially Regulated Genes between Four Sources of hPSC-HPCs versus Somatic HPCs

Table S3. ANOVA Results Comparing CXCR4± hPSC-HPCs with Somatic HPCs

Due to space constraints, Table S1-S3 data omitted from thesis. However, it is present in the open access publication (https://www.cell.com/stem-cell-reports/fulltext/S2213-6711(18)30171-1).

Supplemental experimental procedures

Somatic blood samples
Informed consent was obtained from full-term umbilical CB and MPB donors with protocols approved by the Research Ethics Board at McMaster University. Human BM was purchased from Lonza (Cedarlane, #1M-105). Mononuclear cells (MNCs) were recovered by density gradient centrifugation (Ficoll-Paque, GE Healthcare), and red blood cells were lysed using ammonium chloride (StemCell Technologies). Lineage depletion was performed by magnetic cell separation using a lineage antibody kit (StemCell Technologies, #19309C). Cells were cryopreserved in 10% DMSO in FBS until use.

hPSC culture and differentiation
All experiments were performed using human ESC lines, H9 and CA2, maintained on Matrigel (BD) in mouse embryonic fibroblast-conditioned media (MEF-CM) with 8 ng/mL basic fibroblast growth factor (bFGF), as previously described (Chadwick et al., 2003). In a subset of experiments, CB-induced PSCs (Lee et al., 2014) were additionally tested. Media was changed daily, and cells were passaged as clumps weekly using collagenase IV. Daily morphological evaluation of cells was performed with light microscopy with routine monitoring of pluripotency marker expression (TRA-1-60 and Oct4) by flow cytometry. hPSC-HPCs were produced using the Cytokines and BMP4 protocol, unless stated otherwise, where EB were generated in suspension as previously described (Chadwick et al., 2003). Other hPSC-HPC differentiation methods include OP9 co-culture: MEF-CM-cultured hPSCs were passaged as clumps on to over-confluent OP9 and cultured as previously described (Choi et al., 2011); and EHT: H9 were maintained in mTeSR1 on Matrigel and differentiated as previously described (Lee et al., 2017a), in normoxia (5% CO_2 incubator) or hypoxia (5% O_2/5% CO_2/90% N_2). Upon personal communication with Dr. Lee; ascorbic acid (0.28 mM) and folic acid (0.09 mM) were used at concentrations lower than reported (Lee et al., 2017a). All reagents were purchased from the suppliers listed in each study. Additional compounds were supplemented into Cytokines and BMP4 EB media during hPSC hematopoietic differentiation, as follows:

Small molecules tested for inducing CXCR4

Compound	Common Name	Dose	Treatment Duration	Biological Replicates	Supplier	Reference
ß-estradiol	estrogen	10 nM	48 h, or 72 h	3 per time point	Sigma	(Rodriguez-Lara et al., 2017)

Compound	Common Name	Dose	Treatment Duration	Biological Replicates	Supplier	Reference
16,16-dimethyl-prostaglandin E2	PGE_2	10 μM	2 h, 24 h, 48 h, or 5 d	3 per time point	Cayman Chemical	(Cutler et al., 2013)
fluticasone propionate	flonase	100 nM	24 h	3	Selleck Chemicals	(Guo et al., 2017)
forskolin	forskolin	10 μM	72 h	3	Abcam	(Saxena et al., 2016)

Xenotransplantation

Immunodeficient NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice were bred in a barrier facility and all experimental protocols were approved by the Animal Research Ethics Board of McMaster University. Mouse sex and age were controlled within each experiment, and mice were randomly assigned to experimental groups which included male and female mice. No statistical method was used to predetermine sample size. Mice were sublethally irradiated (single dose of 315 cGy, ^{137}Cs) 24 h before transplant. Cells were transplanted by intra-femoral injection as previously described (Wang, 2005), at doses described in figure legends. At harvest, mice were killed and BM from the injected and contralateral femurs were collected separately. Spleens were separately harvested. Lungs were harvested in a subset of experiments. Cells were recovered by mechanical dissociation in IMDM supplemented with 3% FBS (HyClone, Canada), and 1 mM EDTA (Invitrogen). Red blood cells were lysed using ammonium chloride. BM samples were counted and plated for CFU frequency. BM cells from NSG mice injected with saline were used as negative control for CFU. Phenotyping was analyzed using fluorescence minus one (FMO) gating. Data points were combined from all independent experiments and outliers were not excluded.

Immunophenotyping and cell sorting

For CFU analysis, cells were FACS purified based on CXCR4 and CX3CR1 expression using a MoFlo XDP Cell Sorter (Beckman Coulter). For new microarray gene expression samples, Cytokines & BMP4-generated hPSC-HPC were FACS purified as $CD34^+CD45^+GFP^\pm$ (see Figure 7A), EHT cells as $CD34^+CD43^+$, and CB and BM cells as $CD34^+CD45^+$, using a FACSAria II (BD). For all live staining experiments, $<1x10^6$ cells/200 μL were incubated with antibodies for 30 min at 4°C, and then washed before flow cytometry. 7AAD (Beckman Coulter) or Live/Dead Fixable Violet Dead Cell Stain (Thermo Fisher) was used to exclude nonviable cells. An LSRII Flow Cytometer (BD Biosciences) was used for phenotyping.

Antibody details

Antigen	Reactivity	Conjugated	Dilution	Clone	Supplier
CD34	Human	APC, APC-Cy7, PE	1:100	581	BD Biosciences
CD38	Human	PE	1:100	HB7	BD Biosciences
CD43	Human	FITC, PE	1:100	1G10	BD Biosciences
CD45	Human	FITC, v450	1:100	2D1	BD Biosciences
Cutaneous lymphocyte antigen	Human	FITC	1:100	HECA452	BioLegend
CXCR4	Human	APC, PE	1:50	12G5	BD Biosciences
CX3CR1	Human	PE-Vio770	1:100	2A9-1	Miltenyi
CD45	Mouse	PE-Cy7	1:3000	30-F11	BD Pharmingen

Gene expression profiling

Total RNA was extracted from $0.5\text{-}3.0x10^4$ FACS-purified cells using the RNeasy Micro Kit (Qiagen) following the manufacturer's protocol. RNA was processed using the GeneChip WT

Pico Kit and analyzed with Affymetrix Human Gene 2.0 ST microarray (London Regional Genomics Centre, Ontario, Canada). Gene expression analysis was conducted using Partek Gene Suite (v6.6, Partek Inc). Expression levels of RNA-seq data were obtained from a series matrix sheet in the GEO repository (NCBI). Log2 transformation of RNA-seq data was completed as previously described (Nakamura et al., 2016). For comparison of microarray data to RNA-seq data, the mean probe intensity was used for genes with multiple probes. Datasets were merged by common gene symbols and batch effect was removed using Partek Gene Suite. Genes were considered differentially regulated with fold-change > | 2 | and false detection rate (FDR) p value < 0.05.

Gene expression sample details

Lab	GEO ID	Symbol in study	Samples Used	Sample IDs	Platform	Total Annotated Genes
Bhatia	GSE106721	hexagon	15	All; GSM2849362 to GSM2849376	HG 2.0 ST Array, and HG U133A	34661
Bhatia	GSE92778	circle	6	GSM2437567 to GSM2437572	HG 1.0 ST Array	20796
Bhatia	GSE3823	circle	9	U133A; GSM87705 to GSM87716, GSM87729 to GSM87734	HG U133A	13462
Daley	GSE49938	diamond	17	GSM1210379 to GSM1210384, GSM1210388 to GSM1210392, GSM1210401 to GSM121406	HG U133A Plus2	23520
Daley	GSE83719	triangle	5	All; GSM2214010 to GSM2299187	Illumina NextSeq 500	25855

Gene set enrichment analysis

Global GSEA was performed with default parameters (Subramanian et al., 2005), with gene sets from the Molecular Signatures Database. False discovery rate (FDR) < 0.25 with P < 0.05 was considered significant (Sugimura et al., 2017).

Percent (%) overlap of gene sets used in Pearson correlation and GSEA, with global transcriptome dataset

Figure 7 Group		Molecular Signatures Database Standard Name	Overlap
GPCR	CXCR4	BIOCARTA_CXCR4_PATHWAY	96
		BIOCARTA_AGPCR_PATHWAY	100
	Migration	REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES	74
		GO_G_PROTEIN_COUPLED_CHEMOATTRACTANT_RECEPTOR_ACTIVITY	84
		GO_LEUKOCYTE_CHEMOTAXIS	81
		GO_LEUKOCYTE_MIGRATION	86
Upstream of CXCR4	Hypoxia	BIOCARTA_HIF_PATHWAY	100
		BIOCARTA_VEGF_PATHWAY	93
	Notch	REACTOME_SIGNALING_BY_NOTCH2	83
		REACTOME_SIGNALING_BY_NOTCH	78
	Steroid	GO_GLUCOCORTICOID_RECEPTOR_BINDING	93
	PGE	GO_CELLULAR_RESPONSE_TO_PROSTAGLANDIN_E_STIMULUS	100
		GO_RESPONSE_TO_PROSTAGLANDIN_E	96
Down-stream of CXCR4	Calcium	GO_CALCIUM_MEDIATED_SIGNALING	73
		GO_REGULATION_OF_CALCIUM_ION_TRANSPORT	80
		GO_REGULATION_OF_CYTOSOLIC_CALCIUM_ION_CONCENTRATION	88
	Kinases	GO_ACTIVATION_OF_PROTEIN_KINASE_ACTIVITY	86
		GO_ACTIVATION_OF_MAPK_ACTIVITY	85
	JAK STAT	KEGG_JAK_STAT_SIGNALING_PATHWAY	85

Figure 7 Group		Molecular Signatures Database Standard Name	Overlap
	mTOR	KEGG_MTOR_SIGNALING_PATHWAY	90
HSC Identity	hESC	BENPORATH_ES_1	83
	HSC	EPPERT_HSC_R	76
		JAATINEN_HEMATOPOIETIC_STEM_CELL_UP	70
		JAATINEN_HEMATOPOIETIC_STEM_CELL_DN	84
		GEORGANTAS_HSC_MARKERS	85
		GO_HEMATOPOIETIC_STEM_CELL_PROLIFERATION	77

STRING Analysis

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) analysis was performed with default settings (Szklarczyk et al., 2015), using 120 significantly differentially regulated genes which were upregulated by somatic HPCs compared to hPSC-HPCs, listed in Table S2. Disconnected nodes were removed.

Colony forming unit (CFU) assay

Primary somatic HPCs, hPSC-derived HPCs, and BM transplant samples were plated at $5.0x10^2$-$1.0x10^4$ cells/0.5 mL, $1.0x10^3$-$5.0x10^4$ cells/0.5 mL, and $1.0x10^5$/0.5 mL respectively, in Methocult H4434 (StemCell Technologies). Cells were incubated at 37°C for 14 d and manually scored. Each CFU well represents an independent biological assay, as input cells and MethoCult formulations were individually prepared for testing in single wells. CFU were stained with calcein green (Invitrogen) in Hank's Buffered Salt Solution (HBSS) for 30 min, and imaged with the Operetta High Content Imaging System (PerkinElmer). CFU from transplants were picked by micropipette, genomic DNA was extracted by DNA Micro Kit (Qiagen) following manufacturer's protocol, and analyzed by conventional PCR.

Conventional PCR primer sequences

Human DNA / Gene	Forward (5' – 3')	Reverse (5' – 3')
alpha-satellite, chromosome 17	GGGATAATTTCAGCTGACTAAACAG	TTCCGTTTAGTTAGGTGCAGTTATC
genomic CXCR4 (gen. CXCR4)	GGTGGTCTATGTTGGCGTCT	TCGATGCTGATCCCAATGTA
viral CXCR4 (vir. CXCR4)	TGGAATTTGCCCTTTTTGAG	TTGTCCGTCATGCTTCTCAG
viral eGFP (vir. GFP)	CACATGAAGCAGCACGACTT	TGCTCAGGTAGTGGTTGTCG

Cytospin

CFU were centrifuged onto glass microscope slides using the Shandon Cytospin 3 (Block Scientific, Inc). Differential staining was performed with the Shandon Kwik-Diff Stain Kit (Thermo Scientific).

Lentivirus transgene expression

CXCR4 was subcloned into the pHIV(IRES)EGFP vector (#21373, Addgene). Site directed mutagenesis targeted N123K; termed CXCR4(off). Lentivirus was produced from HEK 293FT cells with 2nd generation pMD2.G and psPAX2 packaging plasmids, and collected after 72 h. The multiplicity of infection (MOI) was calculated by a dilution series on HEK cells, and used at MOI of 100. Experimental cells were incubated with concentrated lentivirus for 48 h in the presence of 8 μg/mL polybrene (Sigma) in Cytokines and BMP4 media (Chadwick et al., 2003).

Transwell assay
1.0-1.5x10^5 cells were seeded in the upper Transwell (EMD Millipore) compartment in 0.4 mL, with 200 ng/mL CXCL12 or CX3CL1 (PeproTech), or 0.01% BSA in 0.6 mL Iscove's Modified Dulbecco's Medium (IMDM, Gibco) in the bottom. Cells in the bottom well were collected after 4 h and counted by flow cytometry. Each Transwell represents an independent biological assay, as input cells and media formulations were individually prepared for testing in single wells.

Calcium flux assay
Cells were adhered to 24-well plates pre-coated with Cell-Tak (Corning) for 3 h. Cells were loaded with 1 μM Fura Red AM (Thermo Fisher) for 60 min at RT, washed twice, and incubated at 60 min RT; all in 25 mM HEPES, 20 mM glucose, in HBSS. Dynamic fluorescent imaging of intracellular calcium concentration at a single focal plane was acquired with an inverted confocal microscope, with an image pair (415 and 485 nm excitation) collected every 2 s for a total of 8 min 40 sec (250 frames). At 50th frame, AMD3100 (10 μM, Mozobil, Genzyme) or DMSO (0.01%) was added, at 100th frame CXCL12 (200 ng/mL) was added, and at 200th frame Ionomycin (10 μM, Sigma-Aldrich) was added as a positive control. Ratiometric analysis of Fura Red intensity over time was quantified using ImageJ.

Chapter 4: Activation of endogenous hematopoietic stem cell regeneration following cytoreductive therapy

Preamble

This Chapter is unpublished data.

Author contributions: J.C.R. performed experiments and received assistance from A.L.B. and J.L.R.G. for single cell RNA-sequencing sample preparation and analysis and received assistance from D.P.P. and D.G. for performing animal work. J.C.R. designed experiments, interpreted data, and wrote the manuscript. M.B. directed the study.

Author contributions in greater detail: J.C.R. wrote the manuscript and performed all experiments. The three of us (J.C.R., A.L.B., and J.L.R.G.) completed the single cell RNA-sequencing library preparation together; all subsequent analyses using R software are my own original work. In 50% of mouse experiments, D.P.P. crushed bones that I dissected, and helped cryopreserve bone marrow for time sensitive experiments. D.G. kept the logistics of bone marrow aspiration organized (depilate mouse knee, weighing and labelling mice) so I can focus on the bone marrow aspiration technique. M.B. designed experiments, interpreted data, directed the study, and provided feedback on the manuscript.

Activation of endogenous hematopoietic regeneration following cytoreductive therapy

Jennifer C. Reid,[1,2] Allison L. Boyd,[1] Juan Luis García-Rodriguez,[1] Deanna P. Porras,[1,2] Diana Golubeva,[1,2] and Mickie Bhatia[1,2,*]

[1]Stem Cell and Cancer Research Institute, Michael G. DeGroote School of Medicine, McMaster University, Hamilton, ON L8N 3Z5, Canada
[2]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON L8N 3Z5, Canada
*Correspondence: mbhatia@mcmaster.ca (M.B.)

Abstract

Radiation and chemotherapy are important cancer treatment modalities, which target excessively proliferating cells. However, these therapies also inflict serious side effects on the hematopoietic system via bone marrow myelosuppression, since hematopoietic stem cells and their progeny produce billions of new healthy blood cells every day that are key to patient survival and health. Radiation and chemotherapy side effects occur in the majority of patients and often require multiple hospital procedures including transfusions, transplants, and may even limit a patient's ability to continue receiving treatment. We investigated multiple modalities of myelosuppression in human-mouse xenograft models and functionally defined biological time points based on progenitor expansion that were conserved across multiple inducers of bone marrow injury. Single cell transcriptional profiling of over 30,000 cells from these biological time points identified conserved markers of macrophage expansion as well as beta-adrenergic receptor signaling. These features were absent in single cell transcriptional profiling of leukemia cells from a relapse patient, but consistently observed in healthy bone marrow as well as a leukemia patient who achieved remission. Deeper single cell transcriptional profiling of the primitive fraction displayed subpopulations where damage persisted whereas other subpopulations were actively proliferating. HRC markers were also associated with genes up-regulated in blastema suggesting a more universal and potentially evolutionarily conserved regenerative program. Although this work is incomplete, this overall approach forms the basis for a campaign to examine drugs and cell transfusions that may activate endogenous regeneration of healthy hematopoiesis when damaged by chemotherapy and radiation.

Introduction

Human tissue regeneration, especially in adulthood, is exceedingly limited and loosely defined across multidisciplinary fields. For example, 'liver regeneration' is more accurately defined as compensatory hyperplasia; proliferation of cells which maintain their differentiated structures and functions (Michalopoulos and DeFrances, 1997). Multiple tissues use compensatory hyperplasia, including β-cells of pancreatic islets (El Ouaamari et al., 2013) and kidneys (Rojas-Canales et al., 2019). Tissue regeneration entails primitive cells which function in repair and replacement of certain cell types upon differentiation, including muscle satellite cells (Wagers and Conboy, 2005) and neural stem cells (Gage and Temple, 2013). These primitive cells have yet to be harnessed to fully replace tissue, however, they hold interminable promise in the field of regenerative medicine. Hematopoietic stem cells (HSC), however, are regenerative in that it is possible to replace all cell types, and is the only proven stem cell therapy.

Chemotherapy and radiation are standard and predominant modalities of most cancer therapy. However, patients receiving this standard of care therapy suffer non-specific off target effects from these cytotoxic treatments, the major one being the destruction of heathy tissue by interfering with normal cell proliferation required for tissue maintenance (Vahdat et al., 2002). Due to high cellular turnover and dependence of cell proliferation, the hematopoietic and immune system are particularly sensitive to cytotoxic injury and therefore patients suffer from low production of white blood cells, red blood cells, and platelets. Indeed, chemotherapy-induced myelosuppression (CIM), acute suppression of BM

and blood cell production, is in fact the most common dose-limiting side effect of most chemotherapy drugs (Wang et al., 2006). HSCs, which functionally represent less than 0.1% of all hematopoietic cells (Delaney et al., 2010), divide asymmetrically to produce progenitor cells which further divide to generate billions of new blood cells every day to sustain tissue homeostasis (Orkin, 2008). Stressed or damaged HSCs repair more slowly than progenitor cells (Milyavsky et al., 2010), will activate programmed cell death, and have impaired repair mechanisms, which can lead to malignant transformation and long term accumulation of mutations (Mohrin et al., 2010). Deterioration of the hematopoietic system leaves patients at risk of infections and poor immune surveillance of tumors, as well as making these patients less eligible for activation-based immunotherapy. Furthermore, CIM may also require withdrawal of chemotherapeutic pressure, which escalates the risk of cancer relapse (Vahdat et al., 2002). Supporting the hematopoietic system using exogenous products such as serial transfusions and transplants provides relief, however, complications, availability, and costs limit broader applications (Crémieux, 2000). The prevailing clinical view is that a linear-log relationship exists between higher chemotherapy dose and even higher tumor cell kill (Vahdat et al., 2002), suggesting efforts to prevent chemotherapy-induced hematopoietic suppression, and therefore discontinuation of treatment, would enable more aggressive and effective anti-cancer therapy. Indeed, granulocyte-colony stimulating factor treatment and prophylaxis antibiotics have historically been used to mitigate neutropenia and infection risks in patients receiving chemotherapy induced by side effects to healthy hematopoietic tissue, but this increases unnecessary treatment and antibiotic resistance, and has not reliably reduced hospital costs nor provided consistent clinical impact (Ozer et al., 2000). As such, an innovative approach to this problem of chemotherapy and radiation induced injury is needed including those that avoid long term effects of HSCs.

The role of cytotoxic damage to HSCs, the apex of the hematopoietic and immune system, is largely unknown. This question has been technically difficult to resolve, as it is well established that HSCs are heterogeneous in their degree of self-renewal, life span, and differentiation capacity (Birbrair and Frenette, 2016). Despite extensive characterization of mature hematopoietic lineages (Orkin, 2008), an incomplete and often contradicting assortment of cell surface markers enrich for long term HSCs, but do not exclusively mark these cells (Taussig et al., 2005). Therefore, functional analyses of self-renewal through in vivo preclinical mouse models remain the gold standard of HSC research. Furthermore, HSCs are tightly regulated, and respond to both cell-intrinsic programs as well as extracellular signals, which include growth factors and hormones (Boyd and Bhatia, 2014; Himburg et al., 2018). These signals are further differentially regulated during steady state homeostasis as well as during times of regeneration after injury induced by cytotoxic therapy. While bulk transcriptomic and functional studies have laid the foundation by which to evaluate the response of HSCs to stress and cytotoxic injury, CIM continues to be an urgent and unmet clinical issue facing many cancer patients.

Our research group recently demonstrated that an anti-leukemia chemotherapy agent, cytarabine (known as AraC), induces a divergent transcriptional signature from healthy HSCs (Boyd et al., 2018); termed hematopoietic regenerating cells (HRC). To broad the application of HRC, we performed serial xenotransplantation and biological progenitor assays to capture lead chemotherapy agents of leukemia and solid tumors, as well as radiation-induction of myelosuppression. We defined functional time point of progenitor expansion following multiple forms of CIM and used single cell transcriptional profiling of

over 30,000 single cells to identify shared programs of regeneration which were validated in healthy and leukemic clinical bone marrow samples, as well as potentially evolutionarily conserved using models organisms of limb regeneration; *Polypterus* and axolotl. Together, these series of experiments suggest there is a previously unappreciated opportunity to support HRCs to improve regeneration which is independent of the cytotoxic agent.

Results

Macroscopic effects of myelosuppression on human hematopoiesis in vivo

Previous work by our research group demonstrated that healthy hematopoietic recovery after cytotoxic treatment occurred more slowly, if at all, than leukemic generating cell (LRC) expansion in human-mouse xenografts when treated with cytarabine (AraC) chemotherapy (Boyd et al., 2018). In order to apply these findings to a broader concept of anti-cancer therapy, we treated naïve (previously untreated) mice with radiation and 5-fluorouracil (5-FU) chemotherapy treatment regimens. We used serial 5-Fluorouracil injections as a model for chemotherapy-induced bone marrow suppression, as it is the drug of choice due to widespread and effective uses in many mono- and combination therapies to treat breast, colorectal, esophageal, stomach, pancreatic, and cervical cancers (Yang et al., 1999), and therefore is classified as an essential medicine by the WHO. 5-FU and AraC are both pyrimidine analogs which interfere with DNA synthesis, and are believed to target rapidly dividing cells. We validated the maximal tolerable dose of radiation (Figure S1A) and 5-FU (Figure S1B). We also validated their myelosuppressive effects via severely hypocellular bone marrow (Figure 1A), transient weight loss (Figure S1C), and 5-FU consistently lead to diarrhea occurrence, all of which are consistent with side effects observed by patients receiving cytotoxic therapy (Vahdat et al., 2002; Wang et al., 2006; Yang et al., 1999).

Since human-mouse xenograft modeling uniquely allows multiple experimental conditions to be tested using one single HSC donor, we used single HSC donor samples for each experiment to conduct controlled transplants. We established homeostatic levels of human HSC engraftment via intravenous injection into immunocompromised NSG mice (Figure 1B). Engraftment was confirmed 7 wk post-transplant with intrafemoral (IF) bone marrow aspirates. This enabled the randomization of xenografts by recipient mouse sex and engraftment level through all treatment groups (Figure 1C, Figure S1D). Murine bone marrow (BM) radiation-induced myelosuppression was more severe than AraC chemotherapy, while 5-FU CIM was more transient than AraC CIM (Figure 1D). This suggested that the kinetics of HRC, reported previously with AraC-based CIM (Boyd et al., 2018), may be different depending on the inducing agent of myelosuppression.

As SCID repopulating cell (SRC) assays are transplantation dependent (Bhatia et al., 1997), they likely introduce technical variables (Sun et al., 2014), that do not apply to CIM in patients. Therefore, we performed repeated BM aspirate sampling in order to mimic clinical standards, using NOD/Scid mice that were preconditioned with homeostatic levels of human HSC engraftment (Figure 1E). Using serial BM aspirates, we assessed the chimerism of human hematopoietic cells (huCD45+msCD45-) following radiation and 5-FU (Figure 1F). Serial BM aspiration of these individual mice following myelosuppression found that healthy HRC respected the boundary of their initial chimerism following radiation and 5-FU treatment and did not exceed it (Figure 1, G-I). Quantitative kinetic modeling indicated that 5-FU treatment,

similar to AraC-based CIM (Boyd et al., 2018), lead to delayed but modest recovery and where radiation lead to complete dysfunctional hematopoiesis and did not recover during this time frame (Figure 1, J-L). Healthy human hematopoiesis showed disciplined patterns of re-growth, with rates of graft doubling time similar to saline-treated controls (Figure 1M). This confirmed our hypothesis that the kinetics of human hematopoietic recovery were correlated to the cytoreduction level of myelosuppression observed in the murine BM. Both 5-FU and AraC-based CIM lead to delayed and modest hematopoietic recovery, while we considered the myeloablative effects of sublethal radiation as our positive control to denote unrecovered dysfunctional hematopoiesis.

Chronology of healthy progenitor regeneration

Having detailed the kinetics of radiation and CIM here, and previously with AraC (Boyd et al., 2018), on the bulk graft of healthy hematopoietic cells, we sought to extend this characterization using the CD34 phenotypic marker for HSPCs. We again initiated myelosuppressive Tx 7 wk post-transplant so that xenografts reached homeostatic chimerism levels prior to analysis (Figure 2A). Mice received BM aspirates the day before myelosuppression, and were randomized by recipient sex and chimerism level. Using flow cytometry to discriminate engrafted human hematopoietic cells (anti-human CD45+), we assessed CD34 expression and progenitor CFU assays of xenografted mice during multiple time points following the completion of myelosuppressive treatments of 5-FU, AraC, and radiation (Figure 2, B and C). As expected, CD34+ composition of control-Tx xenografted mice was unchanged and fluctuated very little between mice (Figure 2D). Human progenitors assessed from 5-FU treated xenografted mice showed a significant and reproducible, yet modest, increase in CD34+ composition within the graft only on day 3 of recovery (Figure 2E), whereas AraC-Tx xenografted mice showed significant increases in CD34+ composition within the graft starting on day 12 of recovery (Figure 2F). Within the limited time points sampled for recovery from radiation, we observed significant suppression of the CD34+ compartment (Figure 2G). These data further support the macroscopic observations of 5-FU leading to a more rapid recovery than radiation (Figure 1).

Since CD34 is a marker of human HSPCs, but serves little known functional significance beyond cell adhesion, we next purified HRC at many regularly spaced intervals of time following myelosuppression to assess quantity and composition of myelo-erythroid progenitors; tested in colony forming unit (CFU) assays (Figure 2B). Similar to the consistency of CD34+ composition in control xenografts, we again observed stable CFU capacity of human cells purified from control-Tx mice (Figure 2H). We continued to use single HSC samples across an entire experiment to control the results, and also confirmed that control samples across multiple donors behaved similarly (Figure S2A). Total progenitor clonogenic output of human progenitors purified from 5-FU-Tx xenografted mice showed a significant increase in CFUs on the third day following treatment completion (Figure 2I). Similar in magnitude but occurring at a later time point, human progenitors assessed from AraC-treated xenografted mice first showed a significant increase in CFUs, starting on the 12^{th} day following treatment completion (Figure 2J). These temporal sequences were similar to the CD34+ composition observed (Figure 2, E and F). Furthermore, these results were also reproduced across multiple HSC donors and independent experiments (Figure S2, B and C). Lastly, we did not observe expansion of the progenitor compartment in radiation-Tx mice

(Figure 2K), which is consistent with the repressed CD34+ compartment observed following radiation-induced myelosuppression (Figure 2G).

Unlike the expansion of leukemic regenerating cells (LRC) we previously reported, which do not produce multi-lineage CFU (Boyd et al., 2018), we observe equal distributions of erythroid (BFU-E), granulocyte (CFU-G), and monocyte (CFU-M) colonies in control-Tx mice engrafted with health HSCs (Figure 2L, Figure S2A). Strikingly, expansion of all myelo-erythroid lineage progenitors was observed on 5-FU day 3 as tested in CFU assays (hematopoietic regenerating progenitor cells, termed "**HRPC^{5FU}**"), with a significant expansion of granulocytic colonies persisting for 6 days (until day 9) of recovery (Figure M). The absolute fold-change of HRPC^{5FU} CFU above control-Tx mice was 4.0-fold. Similar to 5-FU-treated mice, expansion of all myelo-erythroid lineage progenitors was observed on AraC day 12 (termed "**HRPC^{AraC}**"), with a significant expansion of erythroid and granulocytic colonies persisting for 6 days (until day 18). The absolute fold-change of HRPC^{AraC} CFU above control-Tx mice was 5.5-fold. Within the limited time points tested following radiation-induced myelosuppression, significant losses of lineages, most notably granulocytes, were observed Figure 2O, which is supported by the overall loss of CD34+ and functional progenitors (Figure 2, G and K). Previously, we reported the characterization of healthy progenitors on the 9th day following AraC treatment completion (originally termed "HRC") (Boyd et al., 2018), as this time point is of clinical relevance in AML treatment (Aslostovar et al., 2018). In our present study, we refer to the 9th day following AraC treatment completion as "**Nadir^{AraC}**". Similar to the original demonstration of HRC (Boyd et al., 2018), we did not observe progenitor expansion within the graft at the Nadir^{AraC} time point. Given the extremely long myelosuppressive effects of radiation treatment that persist even 9 weeks following radiation-induced myelosuppression (Figure 1), this Tx modality does not appear to capture HRPC within an effective experimental time frame.

Since myelo-erythroid CFU assays to not capture lymphoid tissue progenitors, we characterized the composition of B lymphoid (CD19+) cells in xenografts following myelosuppression using flow cytometry of engrafted human hematopoietic cells (msCD45-huCD45+; Figure S2B). We did not observe temporal changes of the composition of lymphoid cells within the time frame of noted HRPC, which indirectly suggests both types of progenitors are being targeted equally during CIM. Given the identification of the peak HRPC expansion time points for both 5-FU and AraC-based CIM models, we tested cell surface HRC markers we previously defined (Boyd et al., 2018), to initially evaluate whether HRPC expansion was occurring through similar or divergent mechanisms.

Candidate HRC markers of endogenous regeneration

Recently, the prospective isolation of radiation-induced erythroid progenitors was demonstrated in mice (Singbrant et al., 2020), suggesting that phenotypic-based purification of 'stress progenitors' is possible, and this is critically important to determine in the human system as well as identifying supportive cues for regeneration. We first evaluated expression of classical HSC markers such as integrin α6 (known as CD49f; Notta et al., 2011), endothelial protein C receptor (EPCR; Fares et al., 2017), and vascular non-inflammatory molecule 2 (known as GPI-80; Prashad et al., 2015) during recovery from myelosuppression in xenografted mice; which had reached homeostatic chimerism for 7 weeks prior to induction of myelosuppression, and were randomized by recipient sex and engraftment level prior to

treatment (Figure 3A). These three naïve and expansion-based HSC markers were present but were not more preferentially expressed at any of the functionally identified time points (HRPC^{5FU}, HRPC^{AraC}, or Nadir^{AraC}) in the entire graft (Figure 3B) or within CD34+ populations (Figure S3A); suggesting new markers are necessary for prospective identification of HRPC.

We previously reported the population level transcriptome of HRCs compared to therapy-naïve healthy hematopoietic xenografted cells, with the resulting HRC signature highlighting extremely divergent signaling networks as compared to the LRC signature (Boyd et al., 2018). This prompted the investigation of the signaling pathways associated with the HRC signature, with the future goals of promoting healthy recovery and suppressing leukemic expansion in patients. We tested a total of 14 anti-human monoclonal antibodies to cell surface markers identified in the HRC signature (Boyd et al., 2018), represented in the top 67 differentially expressed genes of 302 total genes (Table S1). Including these candidates across multiple flow panels combined with CD45 and CD34, we tested day 3, 6, 9, and 12 following the completion of 5-FU or AraC-based CIM to ensure testing both the HRPC^{5FU} and HRPC^{AraC} time points as well as the original Nadir^{AraC} time point that formed the basis of the HRC signature. Of these 14 candidate HRC markers, ADRB2, ANPEP, CD1c, SIGLEC6, and XCR1 were significantly expressed in human cells from AraC-Tx xenografts compared to control, and of these five, ADRB2 and XCR1 were up-regulated in human cells from 5-FU-Tx xenografts (Figure 3C). Other candidate cell surface markers were not identified, with the limitations of flow cytometry and antibodies directing us to forgo these proteins as prospective markers of HRPC (Figure S3B).

Given the frequency of expression of both ADRB2 and XCR1 at HRPC^{5FU} and HRPC^{AraC} time points, they were tested simultaneously within the CD34+ population (Figure 3D), and all mice were significantly elevated for both candidate markers (Figure 3E). XCR1, also known as lymphotactin receptor, is a g-protein coupled receptor (GPCR) involved in Akt signaling and has established roles in dendritic cells (Saxena et al., 2018). The expansion of dendritic cells is corroborated with additional HRC markers; AXL and SIGLEC6. Beta-2-adrenergic receptor (ADRB2) is also a GPCR, and mediates catecholamine-induced activated cell signaling pathways. In contrast, ANPEP, CD1c, and SIGLEC6 expression was more restricted in AraC-treated xenografts and absent in 5-FU-treated xenografted cells (Figure S3C). These results correlate to the 'cell identity' basis of these markers as opposed to the more identity-independent signaling receptors; ADRB2 and XCR1 which can be expressed by multiple cell types. Initial analysis of subpopulations of hematopoietic cells (Bagger et al., 2018), however, suggested that different subtypes of cells might be contributing to the observed phenotypes of ADRB2 and XCR1 expression (Figure 3F), which demanded a greater resolution to deconvolute the expression patterns of cells in HRPC regeneration which are also dynamically changing in their subpopulation composition.

Transcriptionally distinct regenerating hematopoietic subpopulations

To further characterize the functionally defined HRPC^{5FU} and HRPC^{AraC} samples, and their molecular signaling pathways, we applied single cell RNA-sequencing (scRNA-seq) as a discovery tool (Figure S4A). We also included saline-treated control ('control') and Nadir^{AraC} samples in our analysis to contrast and compare with HRPCs. Comparisons of HRPC transcriptomes versus treatment-naïve cells are expected to form the basis for designing intervention strategies for the activation of endogenous HSPCs. The four samples all

originated from HSC donor 1 (Figure S3A), and were prepared for scRNA-seq through the purification from whole BM of xenografted mice using FACS as both 'whole BM' and 'primitive' (CD34+) samples (Figure S4B). Alignment was performed with CellRanger, and further QC analysis was performed in R using *Seurat* (Stuart et al., 2019) (Table S2). The number of cells which passed QC metrics totaled 24,616 human whole BM cells and 9,635 CD34+ primitive cells, providing substantial population complexity and abundance for analysis. We first integrated the four whole BM samples purified from xenografted mice (Figure 4A), and we assessed the appropriate clustering resolution in R using *Clustree* (Zappia and Oshlack, 2018) (Figure S4C). We then predicted hematopoietic subpopulation cell identities based on established transcriptional markers (Pellin et al., 2019; Villani et al., 2017) and significant entries returned by Enrichr database (Kuleshov et al., 2016) using top differentially expressed genes calculated per cluster (Figure 4C, Figure S4D, Table S3). These cell identities were present in all whole BM samples and observed at frequencies consistent between single cell transcript and protein analysis of the same sample (Figure S4F). Establishing our scRNA-seq data sets are sufficiently complex and representative of the *de novo* samples, we pursued analysis of HRC subpopulations and HRPC candidate markers, as well as discovery of novel markers of stress hematopoiesis.

While myelosuppressive treatments consistently reduced graft chimerism of healthy human hematopoietic cells (Boyd et al., 2018) (Figure 1), the subpopulations of differentiated cells remain unclear. We therefore compared the composition of the four whole BM samples characterized with scRNA-seq (Figure 4C). We did not observe T cells (Figure S4D), which is consistent with human-mouse xenografts using lineage depleted CD34 cells (Doulatov et al., 2012). The composition of each HRC whole BM sample was assessed relative to the control whole BM sample. The composition of HRPC whole BM (both 5-FU and AraC) were strikingly similar to each other, and quite different compared to control whole BM (Figure 4D). We observed expansion of multiple myeloid subpopulations, most notably macrophages and dendritic cells, with the magnitude of expansion larger in $HRPC^{AraC}$ BM than $HRPC^{5FU}$ BM. Modest expansions of neutrophils and monocytes were also observed, at the expense of mature B lymphocytes in both HRPC models. The entire CD34+ compartment and immature B lymphoblasts were severely reduced in $Nadir^{AraC}$ BM cells (Figure 4E), consistent with the reduction of CD34 protein frequency and CFU potential at this time (Figure 2). Macrophages were expanded in $Nadir^{AraC}$ BM, but dendritic cells were suppressed (Figure 4E), suggesting macrophage expansion possibly precedes dendritic cells expansion observed later at peak HRPC recovery (Figure 4D).

To contrast HRC signature genes in a high-throughput measure beyond individual antibody-based protein phenotyping, we computed a module score in R using all available HRC genes simultaneously (Figure 4F, Table S4). We observed macrophages significantly contributed to the HRC score in all samples (Figure 4F), and in equal measures (Figure S5A), suggesting the HRC signature reflects a greater proportion of healthy macrophages. This was supported by the localized expression of lead HRC genes in the macrophage subpopulation (Figure S5B). In contrast, we did not observe up-regulation of classical HSC markers such as CD49f or GPI-80 in HRPC time points in the primitive fraction, while GPI-80 (*VNN2*) appeared up-regulated in monocytes but not macrophages (Figure S5, D and E). To quantify our observations of a potential shared macrophage transcriptional program to the HRC signature, we computed Pearson correlation coefficients to assess the linear dependence of genes on each other at the same time. We compared these coefficients within the

macrophages compartment of all samples combined (Figure 4G). We observed strongly and positively associated correlations between genes in the macrophages, consolidating *in silico* and biological observations that a significant component of healthy regeneration is associated with macrophages. Intriguingly ADRB2 was up-regulated in the CIM BM samples compared to control and restricted in its expression to macrophages in addition to primitive cells (Figure 4H, Figure S5F), which supports its weaker correlation in the HRC signature with classical macrophages (Figure 4G). This observation was significant as it suggests cross talk and stimulation of primitive cells was also likely involved. To validate deeper analysis of the primitive compartment of HRPCs, we first sought translational validation of our observations from CIM-treated xenografts using BM samples from patients undergoing CIM.

Detection of HRPC signaling in leukemia patients receiving HSCT

Publicly available scRNA-seq data sets of healthy human BM, as well as two leukemia patients pre- and post-HSC transplant (HSCT; Figure 5A) were provided by Zheng *et al*. using similar scRNA-seq technology (Zheng et al., 2017). AML Pt.27 was described to experience disease relapse after HSCT, and no graft was observed at relapse. AML Pt.35 was described to achieve remission greater than 3 years (personal communication from Dr. Jerald Radich MD, Clinical Research Division at Fred Hutchinson Cancer Research Center) and the graft at the remission time point was previously described to be comprised of donor cells (Zheng et al., 2017). Both patients received chemotherapy prior to the collection of the second BM sample (personal communication from Dr. Radich). These two patients represent case studies serially monitoring disease relapse or remission, using sophisticated scRNA-seq capture of complex hematopoietic fluctuations over chemotherapy, radiation, and HSCT (Zheng et al., 2017), therefore we evaluated these samples for features of healthy regeneration given that there was a positive and negative outcome case paired with healthy control BM cells. We applied similar QC metrics, visualization, and clustering techniques to interrogate these five samples, representing a total of 24,962 single cells (Figure 5B, Figure S6A). Cell identities were predicted (Figure 5C, Figure S6B) as we previously described for Figure 4; established transcriptional markers (Pellin et al., 2019; Villani et al., 2017) and significant entries returned by Enrichr database (Kuleshov et al., 2016) using top differentially expressed genes calculated per cluster. We observed pronounced expansion of novel cell populations in the relapse sample from patient 27 (Figure 5D), which were mainly distinct transcriptional clusters in the total data set (Figure S6A). Of these two populations, the first, comprising 24% of total cells, was identified as primitive CD34+ expressing cells with no evident lineage specification. Strikingly, the second cluster, comprising 27% of total cells, were identified as 'non-healthy' macrophage cells with shared features of healthy macrophages such as *CD14* and *ITGAM* (CD11b), while also expressing significantly different transcriptional features such as *S100P* and *IL8*; markers of stressed macrophages (Koch et al., 1992; Xia et al., 2018), which were not observed in healthy macrophages (Figure S6, B and C).

Drawing from our observations that the HRC score is predominately influenced by the transcriptome of macrophages from *in vivo* xenograft models of HRPC, we utilized module scoring to identify if HRC markers were particularly enriched in any patient subpopulations or disease states. Indeed, macrophages populations across all patient samples preferentially expressed HRC signature genes (Figure 5E). Of note, the HRC score of the *S100P*+ macrophages in the relapse patient were significantly lower than both classical macrophages

at disease relapse, as well as the same patient prior to relapse (Figure 5F). Furthermore, the classical macrophages were suppressed in terms of proportion and significantly lower in HRC score compared to the sample taken before HSCT and relapse. Interestingly, macrophages pre- and post- HSCT in Pt.35, who ultimately achieved remission, did not experience a proportional increase in *S100P*+ macrophages (Figure 5D), nor a loss of HRC score (Figure 5F). Indeed, the HRC was visually restricted to healthy macrophages and was also highest here, as compared to *S100P*+ macrophages (Figure 5G). To assess potential cooperativity of genes in the HRC signature and their role in macrophages, we computed Pearson correlation coefficients to assess the linear dependence of genes on each other at the same time. We compared these coefficients in both healthy and *S100P*+ macrophages and observed more strongly and positively associated correlations between genes in the healthy macrophages (Figure 5H). This recapitulated our observations of positively correlated HRC genes in healthy macrophages from *in vivo* xenograft models of HRPC (Figure 5G). Of these genes, *HMOX1* showed the biggest difference between strong positive correlation with most genes in healthy macrophages, while reduced positive correlation was observed in *S100P*+ macrophages. This was further corroborated with higher expression of *HMOX1* in healthy macrophages compared to stressed *S100P*+ macrophages (Figure 5I). These data are supportive of the association of macrophage regulation in stress hematopoiesis, and as a distinct phenomenon observed in both xenografts and patients. Encouraged by these similarities, we moved to further characterize the primitive population of CD34 expressing cells, since they are upstream regulators of more mature populations, such as macrophages, and receive microenvironmental cues.

Discovery of conserved HRPC markers of endogenous regeneration

Having observed expansion of the total primitive fraction, phenotypically (Figure 2, E and F), functionally (Figure 2, M and N), as well as transcriptionally using scRNA-seq analysis of whole BM (Figure 6A), it was incumbent upon us to analyze CD34+ subpopulations using scRNA-seq. To discover novel HRPC markers of cells representing functional progenitors, as well as assessing composition and HRC markers, we purified and analyzed CD34+ cells at the two functionally defined HRPC time points as well as Nadir^{AraC} and control cells using scRNA-seq (Figure 6B; also human CD45+ and mouse CD45-), and confirmed the vast majority of cell identities were consistent with *CD34* expressing cells (Figure 6C, Table S5), and we removed a small portion of cells not predicted to be part of the CD34+ compartment prior to further analysis (Figure 6C, Figure S7A). We used *singleR* to predict these cell identities (Aran et al., 2019) given previously described heterogeneity within the human CD34 compartment (Pellin et al., 2019), their description as a continuum when analyzed using single cell RNA-seq, which is largely driven by cell cycle and less by specific transcription factors (Lu et al., 2018). As expected, transcriptional differences between CD34+ HRPC subpopulations were more subtle than between fully mature hematopoietic cells (Figure S7B). Importantly, we did not regress out cell cycle status between cells during normalization, since this appears to drive biologically meaningful insight (Lu et al., 2018). *SingleR* predictions of cell identities found distinct progenitor subpopulations (Figure 6D), which were exceedingly different in samples post-AraC as compared to control and 5-FU treated CD34+ cells (Figure 6E). We observed populations corresponding to HSC identity (shown in brown), common myeloid progenitors (CMP, shown in blue) and granulocyte-monocyte progenitors (GMP, shown in

cyan; more committed than CMPs) to represent a larger proportion of the CD34+ cells in all three post-CIM samples; $HRPC^{5FU}$, $HRPC^{AraC}$, and $Nadir^{AraC}$ (Figure 6F). However, the difference was more pronounced after AraC than 5-FU-based CIM. This observation is supported by the observation of a higher myelo-erythroid CFU clonogenicity of $HRPC^{AraC}$ cells than $HRPC^{5FU}$ cells, since the composition of pro-B cells is much lower in $HRPC^{AraC}$ conditions and B cell colonies are not produced in myelo-erythroid CFU assays (Figure 2L-O).

As expected, CD34+ sorted cells were not particularly enriched in the HRC signature genes (Figure 6G), since this signature is strongly influenced by mature cell types such as macrophages (Figure 4 and 5). This led us to evaluate statistical cross-comparisons between clusters and across samples to discover uniquely regulated genes; signalling independent of cell identity as well as unique cell identity genes which might infer differences within each predicted cluster (Figure 6H). MEPs represent the smallest progenitor compartment analyzed, and yet 89 significantly differentially regulated genes were found unique to this cluster, we ranked these genes based on their fold change between HRPC and control conditions and compared the correlation in expression of the most divergent genes (Figure 6I). We observed three different protein-protein clusters using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) analysis of the most highly expressed genes in the MEP cluster (Figure 6J). However, we did not observe particularly classical MEP-associated genes such as the thrombopoietin receptor (*MPL*, CD135), which is possibly due to transcript drop-out. We similarly analyzed the Pro-B CD34+ cell compartment, which represents the largest progenitor compartment, and 41 significantly differentially regulated genes were found unique to this cluster, which we ranked based on their fold change between HRPC and control conditions and compared the correlation in expression of the most divergent genes (Figure 6K), and used STRING analysis which identified one cluster which was predominately B lineage associated genes (e.g. *CD79A*, *CD79B*, *LCP1*, *VPREB1*; Figure 6L).

We interrogated the CMP compartment more thoroughly in part due to the associations of macrophages in regeneration, but also because of the highly differential clustering of CMPs across samples using *t*-SNE visualization (Figure 6D) as well as their extensive expansion after CIM (Figure 6E). Cluster analysis found 42 differentially expressed genes unique to this cluster, where we extracted the top ten genes and used module scoring to assess the expression of multiple genes simultaneously in each individual CMP cell. We observed a significant progression in the enrichment of these genes from least to most cytopenic BM state after CIM (Figure 6M). Pearson correlation of the most divergent genes between HRPC and control conditions displayed networks of both strongly positive and strongly negative associated genes (Figure 6N), while STRING analysis showed two protein networks were enriched in CMP cells in general. We focussed our attention on *YBX3* expression in CMP cells, due to its positive and negative associations with other CMP genes as well as its highly variable expression in CMP cells (Figure 6P). *YBX3* is transcription factor and repressor of the granulocyte-macrophage colony-stimulating factor (GM-CSF) promoter, GM-CSF promotes stimulation of granulocyte and macrophage differentiation, as well the association of *YBX3* with quiescence and higher expression in G_0 mouse hematopoietic progenitor cells (Lu et al., 2016). *YBX3* expression was significantly higher in post-AraC CIM conditions (Figure 6Q), which appears counter-intuitive in both differentiation and quiescence contexts and remains to be further characterized.

Finally, in all progenitor subpopulations, we observed striking diametrically opposed expression patterns of *MKI67* (Ki-67, marker of proliferation) and *HMGB1* (damage-

associated protein), suggesting each progenitor cluster contains further subpopulations of active and damaged progenitor cells (Figure 6R, Figure S7C). HMGB1 is associated with immune suppression and stressed cells (Parker et al., 2014). Despite the anti-apoptotic and stress signals common to multiple clusters, *MKI67* expression suggests some progenitor subpopulations are undergoing active expansion. Indeed, CMPs and MEPs have been described to represent a continuum, which is largely driven by cell cycle and less by specific transcription factors (Lu et al., 2018). The molecular and biological characterization of differences between these subpopulations of CD34+ HRPCs remains to be further characterized.

Evolutionarily conserved early response of regenerative signaling

In searching for an HRPC signature independent of the agent of injury, we also questioned whether there were any evolutionarily conserved regenerative progenitor cell genes. In considering evolutionary tissue regeneration, the salamander known as the axolotl (*Ambystoma mexicanum*), is perhaps the most well-known representation of complex tissue regeneration (Bryant et al., 2017a; Echeverri and Tanaka, 2002). Axolotl limb regeneration can be exhausted through repeat amputation (Bryant et al., 2017b), while similarly HSCs can be exhausted with repeated 5-FU administration (He et al., 2017). To diversify further, we included expression analysis of the more distally related *Polypterus* fish which can regenerate fins after amputation and have been serially transcriptionally profiled based on functional time points established by blastema morphology; a mass of proliferative cells of many lineages and also thought to contain primitive progenitor cells types (Darnet et al., 2019) (Figure 7A). We observed the majority of genes present in the HRC signature were also conserved across human-*Polypterus* species boundaries (Figure 7B). Using available population-level transcriptional analysis, we found recurrent genes highly expressed in early stages of recovery which were then temporally down-regulated (Figure 7C). Macrophage-associated genes such as *ANPEP*, *CPVL*, and *HMOX1* were highest at early blastema formation, as well as *ADRB2* expression; consistent with up-regulation of these genes in our HRC signature (Figure 7D) as well as protein expression at functionally defined time points (Figure 3). Similarly, axolotl regenerative states are based on blastema morphology (Darnet et al., 2019) (Figure 7E) and the majority of genes present in the HRC signature were also conserved across human-axolotl species boundaries (Figure 7F). We computed a list of these conserved genes ranked by highest differential between control (no amputation) and blastema tissue (Figure 7G). We expected and observed SALL3 expression in both blastema data sets, given its association in limb patterning and regeneration (Stewart et al., 2013). Macrophage-associated genes were again up-regulated upon blastema formation; consistent with up-regulation of these genes in our HRC signature (Figure 7H). We were encouraged to see similar genes represented in the top of both axolotl and *Polypterus* regeneration, which also are strongly associated with our HRC model (Figure 7I). While this *in silico* analysis is also presently correlative, we suggest it may narrow down HRC targets for more universal applications to promote regeneration than presently investigated.

Discussion

The evolutionary origins of basic cellular fate decisions have drawn comparisons between humans and axolotls, such as the role of Oct4 in reprogramming cells to pluripotency (Tapia et al., 2012). Axolotls can regenerate entire multi-tissue body parts such as their limbs and tails without scarring (Darnet et al., 2019; Echeverri and Tanaka, 2002), and this process was originally defined morphologically since their genome is exceedingly large and was difficult to map until only recently (Nowoshilow et al., 2018). Despite the extensive advances in transcriptional profiling of the limb blastema, the ensuing long lists of candidate genes remain largely uncharacterized (Bryant et al., 2017a; Gerber et al., 2018; Stewart et al., 2013; Voss et al., 2015). Here, we compared genes across species which are enriched in repairing tissue. Up-regulation of macrophage genes (e.g. ANPEP, CPVL, and HMOX1), were remarkably conserved across human xenograft models, human HSCT recipients, and multiple blastema.

The role of macrophages in wound repair is well established, as well as the clinical use of granulocyte-macrophage colony stimulating factor (GM-CSF) for patients experiencing CIM (Antman et al., 1988). However, macrophages are also involved in persistent wounds such as diabetic ulcers (Theocharidis et al., 2018), therefore the nature and role of these recurrent cells remains of interest. Macrophage transplantation during acute myelosuppression is of interest, especially since our model suggests macrophages at time of HRPC don't appear transcriptionally distinct from control homeostatic macrophages, suggesting infusion of naïve macrophages may be effective. Infusion of homeostatic macrophages has been shown to improve the recovery of hemolysis in liver protein knock-out mouse models (Kim et al., 2018), and are known to be recruited to sites of radiation-induced injury (Meziani et al., 2018). Developing more insight into the role and nature of these macrophages, as well as potential crosstalk or shared signaling pathways with HSPCs during hematopoietic regeneration remains of interest. In this study, we also present protein targets associated with healthy blood regeneration following chemotherapy and radiation-induced cytotoxicity and evaluated expression of these genes using clinical samples from patients receiving HSCT. Future partnerships and work will determine the efficacy of available chemical agonists (new or available as off patent repurposed compounds) to test targeting of HRC towards activating endogenous HSCs of patients experiencing hematopoietic and immune suppression following therapy and injury, and diseases such as leukemia, HIV, and auto-immune disorders.

Materials and Methods

HSC samples

Informed consent was obtained from full-term umbilical CB donors with protocols approved by the Research Ethics Board at McMaster University. Mononuclear cells (MNCs) were recovered by density gradient centrifugation (Ficoll-Paque, GE Healthcare), and red blood cells were lysed using ammonium chloride (StemCell Technologies). Lineage depletion was performed by magnetic cell separation using a lineage antibody kit (StemCell Technologies, #19309C). Cells were cryopreserved in liquid nitrogen in 10% DMSO in FBS until use.

Flow Cytometry

Cells were stained using antibodies in the Table below. Live cells were stained for 30 min with 1:100 Human TruStain FcX (Fc Receptor Blocking Solution, Biolegend, #422301) and 1:100 Fc block (rat anti-mouse CD16/CD32, clone 2.4G2, #553142) at 4°C in PEF (PBS with 3% FBS (Performance, Wisent, Canada) and 1 mM EDTA (Invitrogen)), washed, and analyzed with an LSRII or Aria II flow cytometers (BD Biosciences). 7AAD 1:50 (Beckman Coulter) was used to exclude nonviable cells. Fluorescence minus one (FMO; if applicable, secondary antibody but no primary antibody), controls were used to set gates in FlowJo (TreeStar, v10).

Antigen	Clone	Fluorochrome	Dilution	Catalogue	Supplier
ADRB2	polyclonal	Alexa Fluor 647	1:100	bs-0947R-A647	Bioss Antibodies
AXL	108724	PE	1:100	FAB154P	R&D Systems
CD1C	F10/21A3	PE	1:100	564900	BD Biosciences
CD13 (ANPEP)	WM15	FITC	1:100	11-0138-42	ThermoFisher Scientific
CD3	UCHT1	PE	1:100	IM1282U	Beckman Coulter
CD19	HIB19	APC	1:100	555415	BD Biosciences
CD235a (Glycophorin A)	11E4B-7-6	PE	1:100	COIM2210	Beckman Coulter
CD327 (SIGLEC6)	767329	FITC	1:100	FAB2859G	R&D Systems
CD33	P67.6	PE	1:100	347787	BD Biosciences
CD34	581	APC-Cy7	1:500	343514	BioLegend
CD41b (ITGA2B)	HIP2	FITC	1:100	555469	BD Biosciences
CD45	2D1	v450	1:100	642275	BD Biosciences
CD49f	GoH3	FITC	1:100	561893	BD Biosciences
CD71	L01.1	APC	1:100	341028	BD Biosciences
CDH1 (E-Cadherin)	180224	PE	1:100	FAB18381P-025	R&D Systems
ENPP1	polyclonal	APC	1:100	FAB6136A	R&D Systems
EPCR (CD201)	RCR-252	APC	1:100	563622	BD Biosciences
FPR3	374822	APC	1:100	FAB3896A	R&D Systems
GPI-80 (VNN2)	3H9	PE	1:100	D087-5	MBL International
HHLA2	907812	Alexa Fluor 647	1:100	FAB80841R-025	R&D Systems
SLC12A5	905705	Alexa Fluor 488	1:100	FAB8369G-100UG	R&D Systems
SLC1A6	polyclonal	purified	1:100	ls-c179249-50	LifeSpan Biosciences
TIMP3	277128	PE	1:100	IC9731P	R&D Systems
XCR1	1097A	FITC	1:100	FAB8571G-100	R&D Systems

Xenotransplantation

Immunodeficient NOD.Cg-Prkdc^{scid} (NOD/SCID) and NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice were bred in a barrier facility and all experimental protocols were approved by the Animal Research Ethics Board of McMaster University. Mouse sex and age were controlled

within each experiment, and mice were randomly assigned to experimental groups which included male and female mice. No statistical method was used to predetermine sample size. NOD/SCID and NSG mice were sublethally irradiated (single dose of 325 and 315 cGy, ^{137}Cs, respectively) 24 h before transplant. Cells were transplanted by intravenous tail vein injection at 0.9-1.5x10^5 CD34+ cells per mouse. Mice were assessed for human chimerism using intra-femoral injection technique as previously described (Reid et al., 2018). Five daily consecutive intraperitoneal (IP) injections of 30 mg/kg 5-FU (Cayman Chemicals, #14416) in 0.9% saline were determined to be the maximal tolerable dose which induced 20% mortality. Five daily consecutive subcutaneous (SQ) injections of 50 mg/kg AraC (Sigma, #C1768-1G) were determined as previously described by our group (Boyd et al., 2018).

Xenograft BM processing

At harvest, mice were killed and BM from the femurs, iliac crests, and tibiae were collected using sterile technique and pooled as one sample, and spine BM was collected separately. Cells were recovered by mechanical dissociation in IMDM supplemented with 3% FBS (Performance, Wisent, Canada), and 1 mM EDTA (Invitrogen), with media kept on ice. Red blood cells were lysed using ammonium chloride at 4°C for 10 minutes. Data points were combined from all independent experiments and outliers were not excluded.

Sample processing for single cell RNA-seq

Harvests were performed quickly and on ice and spine BM MNCa were cryopreserved within an hour of mouse death as described in *Xenotransplantation* section. Cells were stored in liquid nitrogen in 10% DMSO in FBS until use. The leg BM of each of these xenografted mice were assessed for human chimerism, HRC marker expression, and hematopoietic cell lineages by flow cytometry; and tested for human CFU capacity. Two BM samples were thawed and processed for scRNA-seq per day on two back-to-back days. Cells were thawed using the Lonza protocol, and column depletion was performed on each sample with myelin microbeads, mouse CD45 microbeads, and Ter119 microbeads, Fc block, and stained on ice for 20 minutes (LS columns; Miltenyi). Cells were resuspended in PEF kept on ice. Samples were stained (as described in *Flow cytometry* section) with anti-human CD45-v450, anti-human CD34-APC anti-mouse CD45- Alexa Fluor 488, anti-mouse Ter119-Alexa Fluor 488, and sorted in under an hour with an Aria II flow cytometer (BD Biosciences); as 'whole BM' (human CD45+, mouse CD45-, mouse Ter119-, 7AAD-) and as 'CD34+' (human CD45+, human CD34+, mouse CD45-, mouse Ter119-, 7AAD-). Cell count and viability were quantified with a Countess (Invitrogen), and encapsulated into single droplets using Chromium Controller and Chromium Single Cell 3' GEM kit (10x Genomics, (#1000077) and cDNA was prepared the same day as the sort, according to manufacturer's protocols. The GEM volume for AraCd12_CD34 was approximately a third of the expected volume, we suspect a clog occurred, so cell number was lower than expected in this data set. All eight samples were processed for library construction together using a multi-channel pipette, with the Chromium Single Cell 3' Library kit (10X Genomics, #1000078), according to manufacturer's protocols. cDNA and library DNA were quantified using 2100 Bioanalyzer (High Sensitivity DNA Assay Agilent Technologies) at the Farncombe Genomics Facility (McMaster University).

Single cell RNA-seq and analysis

Libraries were sequenced on a NovaSeq SP flowcell (100 cycles) by the Sequencing Facility at The Centre for Applied Genomics (The Hospital for Sick Children, Toronto). Sequencing fastq files of the libraries were processed by CellRanger *count* (10x Genomics, v3.1) to obtain bam alignment files using GRCh38-3.0.0 as a reference. We obtained an average of 103±16 million sequencing reads per sample (range: 67-119 million). Details of individual sample QC are shown in Table S3. Normalization using *Seurat* package (v3.1.2) (Stuart et al., 2019) in R removed cells with >4100 unique transcripts, <200 unique transcripts, and >50% mitochondrial genes; we then removed the effects of the abundance of mitochondrial genes by regression using *sctransform*. Cells that passed QC criteria were then used for analysis. Four 'whole BM' samples (mouse CD45-, Ter119-, human CD45+) were combined using reference-based integration in *Seurat* and visualized using UMAP (30 dimensions). Four CD34+ sorted samples (mouse CD45-, Ter119-, human CD45+, human CD34+) were combined using reference-based integration in *Seurat* and visualized using *t*-SNE in *SingleR*. Importantly, all single cell samples originated from the same HSC donor, and CD34+ samples were sorted from the same whole BM samples presented in this study. After a range of cluster resolution parameters were tested using *clustree* in R (Zappia and Oshlack, 2018), 1.6 and 0.5 were selected as the resolution for 24,616 'whole BM' cells and 9,635 'CD34+ sorted' cells, respectively, because resulting clusters captured major, statistically validated transcriptional distinctions of interest, while avoiding subdivisions of relatively uniform parts of the data. Marker genes for the individual clusters were identified using *findconservedmarkers* function in *Seurat* and used based on criteria of minimum log fold change (log.FC) ≥ 1 and false detection rate (FDR)-corrected P < 0.00001. For 'whole BM' samples, we delineated hematopoietic subpopulation cell identities based on established transcriptional markers (Pellin et al., 2019; Villani et al., 2017) and significant entries returned by Enrichr database (Kuleshov et al., 2016) using top differentially expressed genes calculated per cluster. For 'CD34+ sorted' cells we delineated hematopoietic subpopulation cell identities based on unbiased cell identity predictions using *SingleR* in R (v3.1.2) (Aran et al., 2019), these cells are displayed using *t*-SNE since UMAP is not yet available in this package. HRC score calculations of single cells using multiple genes was completed using *addmoldulescore* with *Seurat* objects. To visualize the data, *t*-SNE, PCA, UMAP, violin plots, and heatmaps were created using *ggplot2* (v3.2.1) or *Seurat*. Pearson correlation matrix of significantly differentially expressed genes was computed with *corrplot* using the *Hmisc* package in R, and genes are ranked by angular order of eigenvectors (*AOE*). Values represented in heatmaps are normalized UMI log counts, except in Figure 6R and Figure S7C (bottom row). Here, two genes, *MKI67* and *HMGB1*, were linearly scaled and mapped from 1 (highest *MKI67* expression) to 0 (lowest *MKI67* expression), and -1 (highest *HMGB1* expression) to 0 (lowest *HMGB1* expression). Then these two scores were combined for each individual cell to produce one single score where 1 represents highest *MKI67* expression and zero *HMGB1* expression; 0 represents equivalent expression between both genes, and -1 represents highest *HMGB1* expression and zero *MKI67* expression.

Publicly accessible data sets

Publicly accessible data sets were downloaded for scRNA-seq human BM samples and AML patients (Zheng et al., 2017), and RNA-seq axolotl and *Polypterus* gene expression (Darnet et al., 2019). STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) analysis was

performed with default settings (Szklarczyk et al., 2015), using significantly differentially regulated genes, and disconnected nodes were removed.

Quantification and Statistical Analysis
Quantification methods were described in the figure legends or specific method details. Number of replicates used were indicated in the individual figure legends. Results were expressed as the means and error bars represent SD unless otherwise indicated. Difference between two groups of samples were examined by unpaired two-tailed Student's t test, and more than two groups using one-way ANOVA, using Prism 7 (GraphPad). In all cases, the data was assumed to meet t test requirements (normal distribution and similar variance) and differences were considered to be statistically significant when $P < 0.05$.

Data and Software Availability
R scripts used for the scRNA-seq analyses are available upon request.

Acknowledgements

We wish to thank the McMaster University Medical Centre Labour and Delivery nurses and parents who were involved in donating umbilical cord blood for this study, as well as Z. Shapovalova and M. Subapanditha for assistance with FACS, and E. Broder and M.-P. Mathieu for assistance with animal husbandry and CB processing. **Funding:** J.C.R. was supported by the Canadian Institutes of Health Research Doctoral Award, an Ontario Graduate Fellowship, and Michael G. DeGroote Doctoral Scholarship.

Figures and Tables

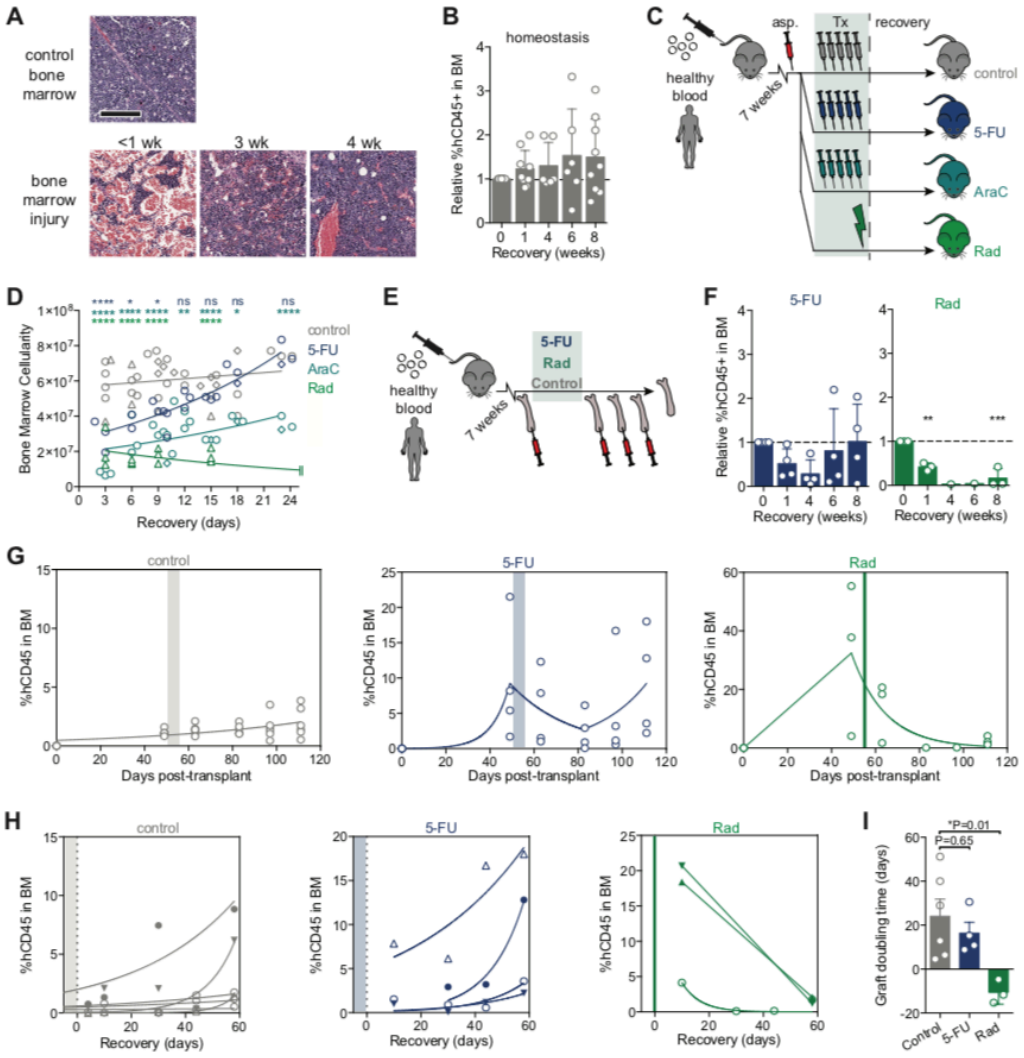


Figure 1. Acute myelosuppression impairs long term hematopoiesis

(**A**) BM morphology of adult mice (H&E stain). Scale bar is 200 μm. (**B**) Healthy chimerism relative to level at 7 wk post-transplant. (**C**) Schematic of myelosuppressive treatments in xenografted mice. (**D**) BM counts of xenografted mice harvested at indicated time points. Each dot is one mouse. Two-way ANOVA. (**E**) Schematic of cytoreductive treatment and aspirate follow-up. (**F**) Healthy chimerism (%hCD45+) relative to level at day -6 (1 day prior to cytoreductive treatment start) of xenografts sampled over time. Each dot is one mouse. One-way ANOVA. (**G**) Healthy chimerism (%hCD45+) in BM over time, sampled by repeated aspirate, since the date of transplant. Mean chimerism is shown with a line, each dot is one mouse. (**H**) Separate depictions of individual xenografted mice assessed for exponential growth of human graft. Each symbol is a different mouse, serially aspirated over time. (**I**) Doubling time of human graft (as calculated in **H**), each dot is one mouse. One-way ANOVA.

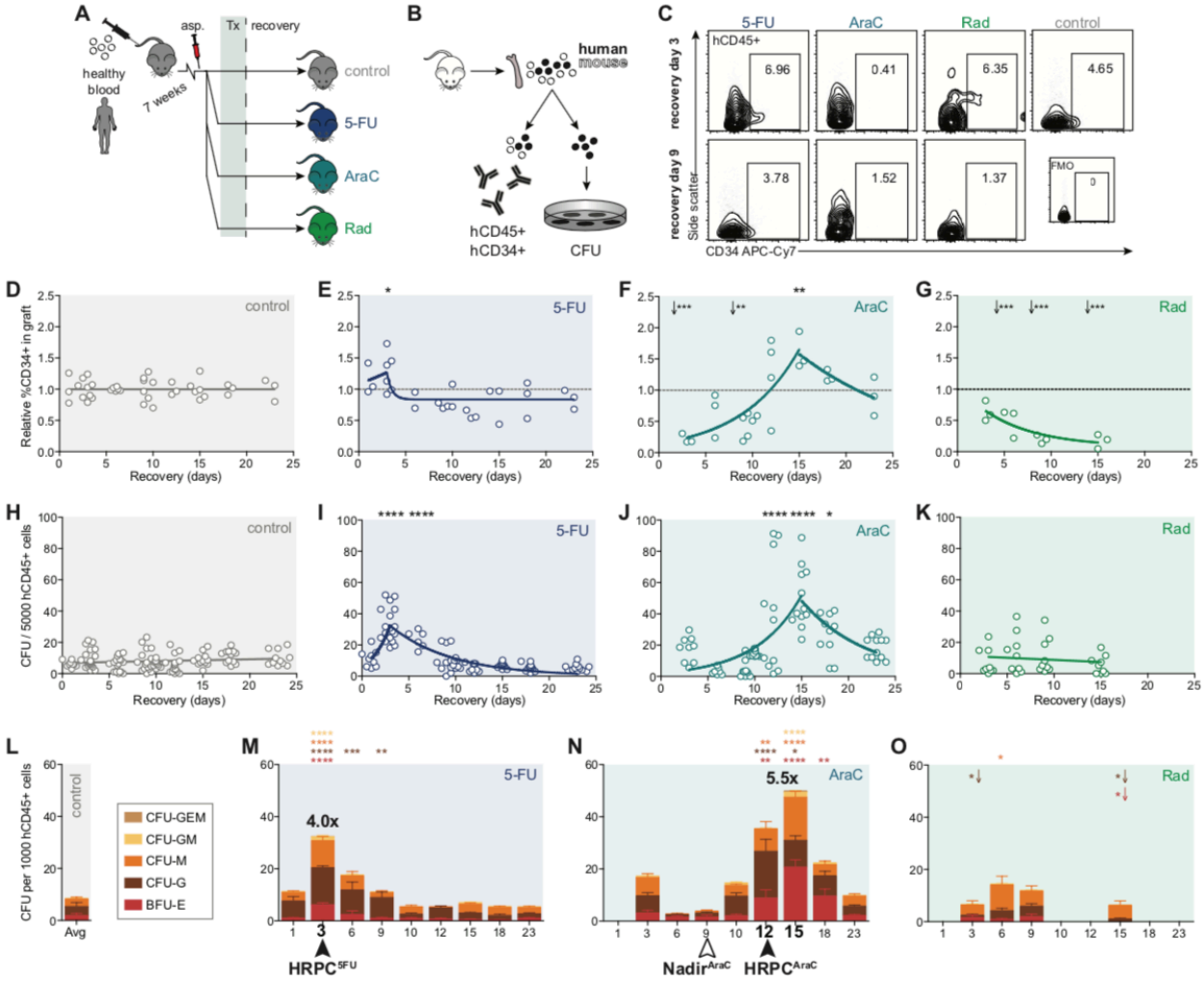


Figure 2. Hematopoietic regenerating progenitor cell expansion following CIM

(**A**) Schematic of myelosuppressive treatments in xenografted mice. (**B**) Schematic of human cells isolated from xenografts after Tx and assessed for CD34 phenotype and clonogenic potential (CFU). (**C**) Flow cytometry of CD34+ composition of representative mice shown on day 3 and 9 of recovery from myelosuppressive treatment. FMO, fluorescence minus one. (**D-G**) CD34+ composition within the human graft of individual mice harvested at the indicated time points. Each dot is one mouse. Two-way ANOVA vs. control mice on same day. (**H-K**) Total CFU counted per 5000 human cells seeded in methylcellulose, of human cells purified from individual xenografted mice harvested at the indicated time points. Each dot is one independent CFU well, n=3 wells per mouse. Two-way ANOVA vs. control mice on same day. (**L**) Average CFU lineage composition of human cells purified from individual control-Tx xenografted mice. (**M**) CFU lineage composition of human cells purified from individual 5-FU-Tx xenografted mice. HRPC peak is 3rd day of recovery (black arrow, $HRPC^{5FU}$). (**N**) CFU lineage composition of human cells purified from individual AraC-Tx xenografted mice, first HRPC day is 12th day of recovery (black arrow, $HRPC^{AraC}$). HRC time point previously reported (Boyd et al., 2018), is depicted as the nadir on day 9 of recovery (white arrow, $Nadir^{AraC}$). For all panels, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

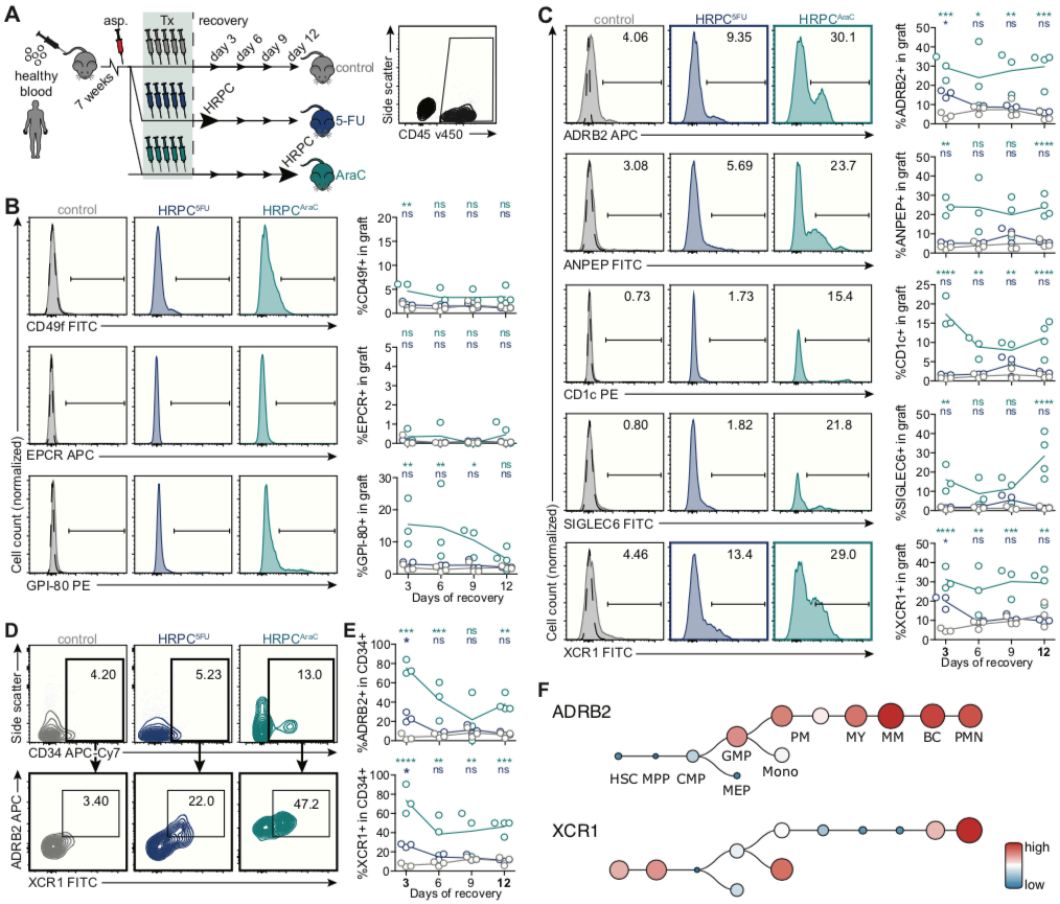


Figure 3. Identification of candidate HRC markers for prospective isolation

(**A**) Schematic of xenotransplantation, CIM Tx *in vivo*, and harvest days which include functionally defined HRPC time points, all samples were first gated by human CD45 positivity (hCD45+). (**B**) Evaluation of classical HSC cell surface markers on whole BM (hCD45+) from xenografted mice. Each dot is one mouse. Two-way ANOVA vs. control Tx. (**C**) Evaluation of candidate HRC cell surface markers on whole BM from xenografted mice. Each dot is one mouse. Two-way ANOVA vs. control Tx. (**D**) Representative flow cytometric plots of HRPC time points and co-expression of ADRB2 and XCR1 on whole BM (hCD45+) from xenografted mice. (**E**) Quantitation of candidate HRC markers within CD34+CD45+ primitive population of each xenograft. Each dot is one mouse. Two-way ANOVA vs. control Tx. (**F**) Hierarchical differentiation tree representing relative expression within discrete hematopoietic subpopulations (adapted from Bloodspot.com). Abbreviations and cell identities further described in Figure S4D. For all panels, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

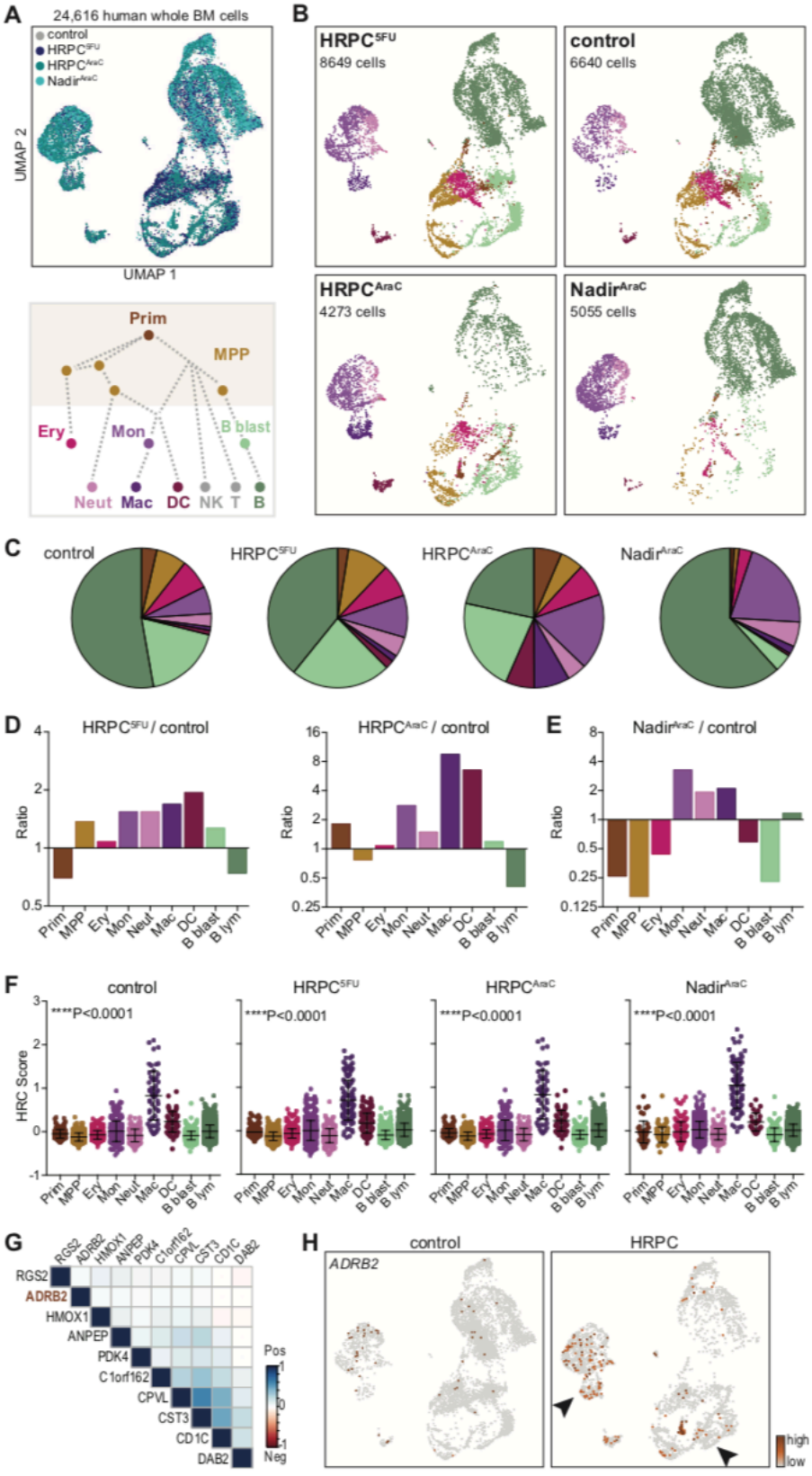


Figure legend on next page

Figure 4. Single cell RNA-seq analysis of human HRC

(**A**) After QC was performed, a total of 24,616 human whole BM cells from 4 samples were integrated and visualized using UMAP in R. (**B**) Hematopoietic cell identity predictions (legend in grey box) of all whole BM single cells from 4 samples are shown. (**C**) Proportion of each cell identity within each sample (legend in grey box). (**D**) Ratio of cell proportions within HRPC whole BM samples compared to the control whole BM sample shown in Figure 4C (legend in grey box). (**E**) Ratio of cell proportions within Nadir^{AraC} whole BM samples compared to the control whole BM sample shown in Figure 4C (legend in grey box). (**F**) Module score of the HRC signature genes (Table S4) for each cell within each of the four whole BM samples, presented grouped by cell identity. Kruskal-Wallis one-way ANOVA. (**G**) Pearson correlation matrix of top differentially expressed HRC gene present in the dataset, establishing dependence between multiple genes at the same time. Positively correlated genes are blue, negatively correlated genes are red; the genes are ranked by the angular order of the eigenvectors. (**H**) *ADRB2* expression visualized using UMAP. HRPC refers to HRPC^{AraC} whole BM sample.

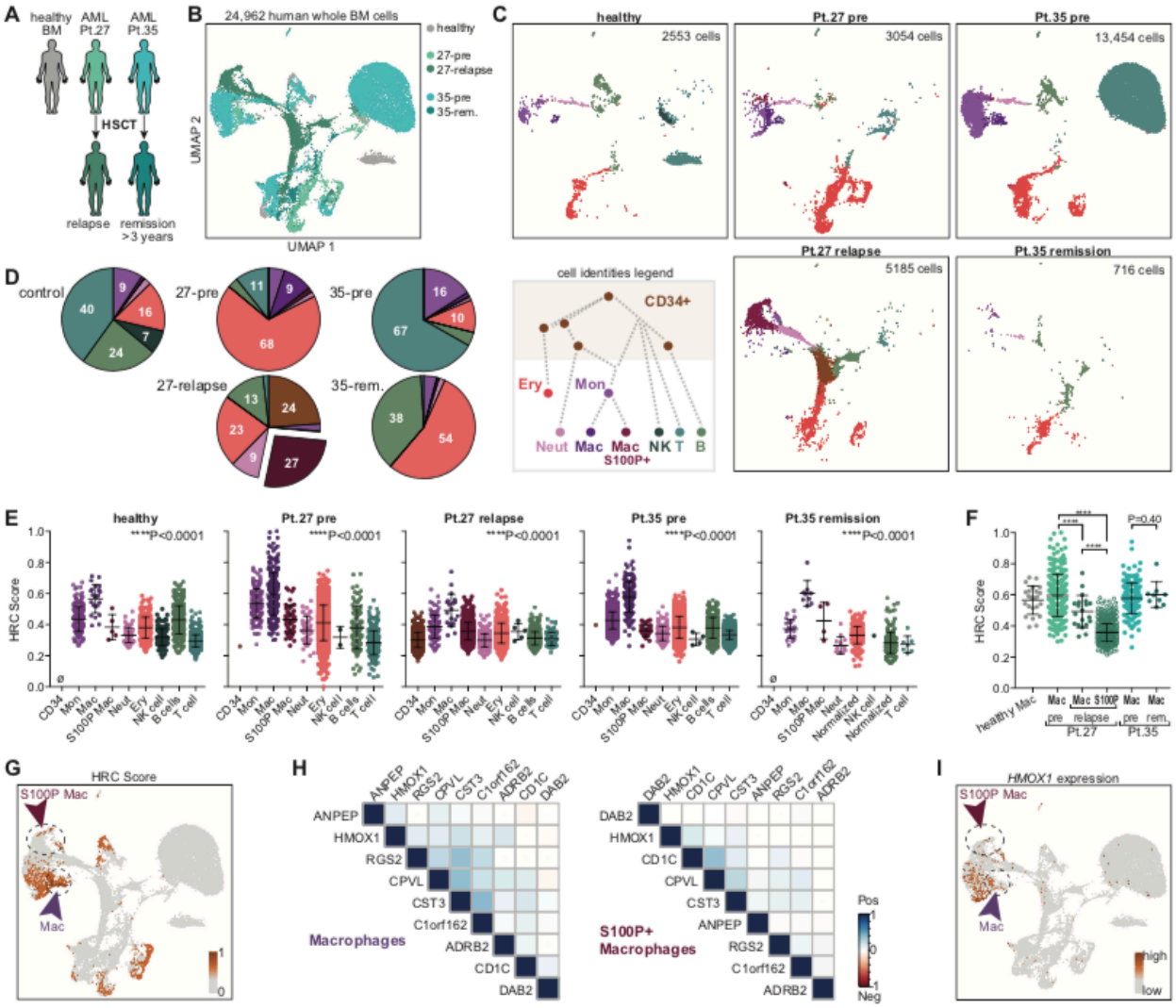


Figure 5. Single cell RNA-seq analysis of human HSC transplant recipients

(**A**) Schematic detailing sample identity and patient outcome. Paired samples from the same patient (Pt) are represented by a darker shade. (**B**) After QC was performed, a total of 24,962 human whole BM cells from 5 samples were integrated visualized using UMAP. (**C**) Hematopoietic cell identity predictions (legend in grey box) of all whole BM single cells from 5 samples are shown visualized using UMAP. (**D**) Proportion of each cell identity within each sample (legend in grey box). (**E**) Module score of the HRC signature genes (Table S4) for each cell within each of the 5 samples, presented grouped by cell identity (each dot is one cell), linearly scaled and mapped between 1 (highest) and 0 (lowest). Kruskal-Wallis one-way ANOVA. Ø indicates no cells in this cluster. (**F**) Module score of the HRC signature genes (Table S4) for macrophages within the 5 samples (each dot is one cell). One-way ANOVA. (**G**) HRC module score per cell, visualized using UMAP. (**H**) Pearson correlation matrix of top differentially expressed HRC genes present in the dataset, establishing dependence between multiple genes. Positively correlated genes are blue, negatively correlated genes are red; genes are ranked by angular order of eigenvectors. (**I**) Highly positively correlated HRC marker *HMOX1* expression visualized using UMAP.

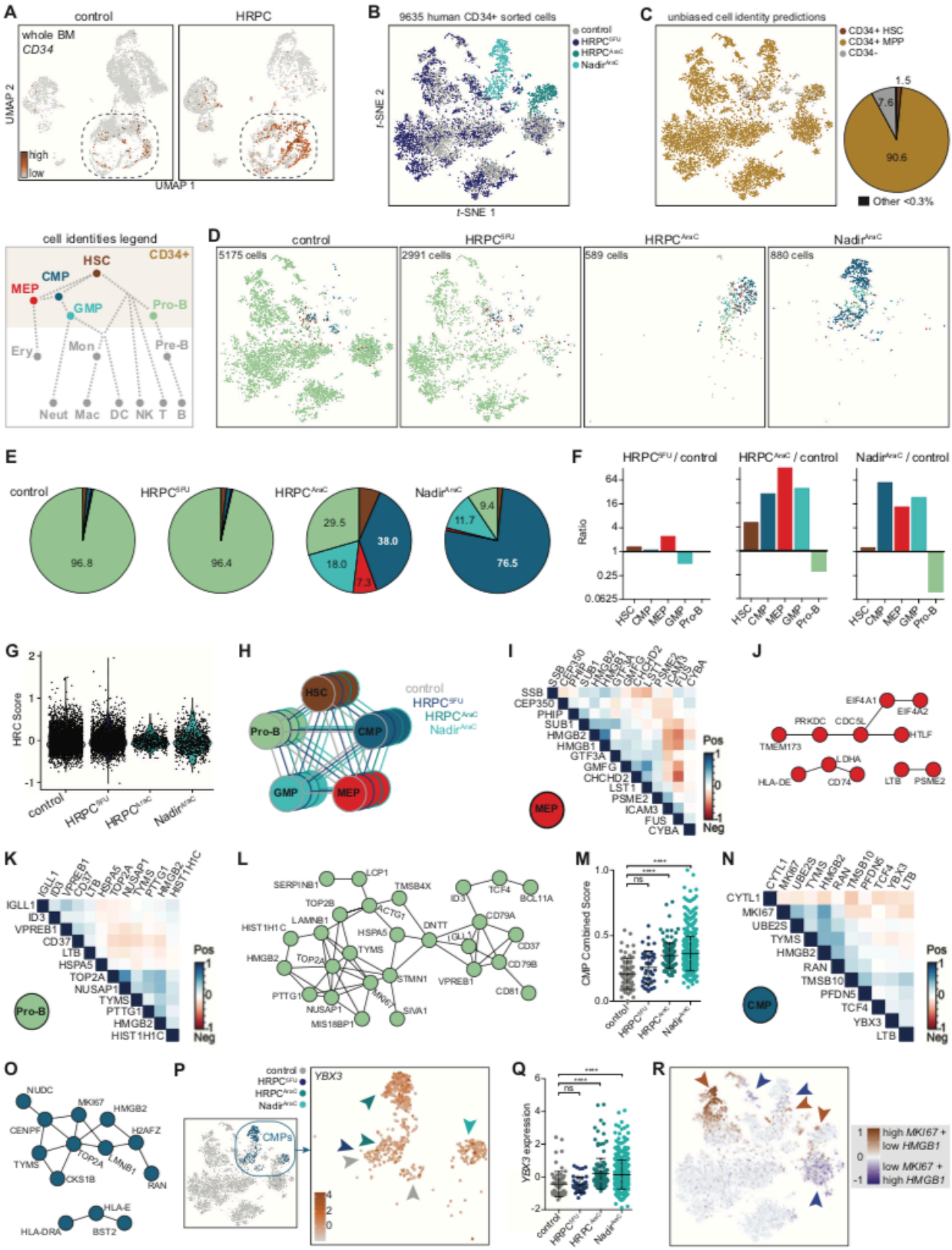


Figure legend on next page

Figure 6. Single cell RNA-seq analysis of human primitive HRPC

(**A**) CD34 expression of whole BM samples purified from HRPC xenograft model, visualized using UMAP. HRPC refers to $HRPC^{AraC}$ whole BM sample. (**B**) After QC was performed, a total of 9,635 human CD34+ sorted cells from the HRPC xenograft model were integrated and visualized using *t*-SNE, which is compatible with *SingleR*. (**C**) *SingleR* predicted cell identities of human CD34+ sorted HRPCs grouped by CD34 expression, visualized using *t*-SNE. CD34+ population: CMP (common myeloid progenitor), MEP (megakaryocyte–erythroid progenitor), GMP (granulocyte-monocyte progenitor), Pro-B cell CD34+, and HSC. CD34- cell identities listed in Table S5. (**D**) Hematopoietic cell identities, based on *SingleR* predictions (legend in grey box), from all 4 CD34+ single cell samples are visualized using *t*-SNE. (**E**) Proportion of each cell identity within each sample (legend in grey box). (**F**) Ratio of cell proportions within CD34+ sorted HRPC xenograft samples compared to the control sample shown in Figure 6E (legend in grey box). (**G**) Module score of the HRC signature genes for each cell within each of the 4 samples. Each dot is one cell. (**H**) Schematic representing statistical cross-comparisons between clusters and across samples to discover uniquely regulated genes; signalling independent of cell identity as well as unique cell identity genes which might infer differences within each predicted cluster. (**I**) Pearson correlation matrix of significantly differentially expressed genes present in MEP cluster, establishing dependence between multiple genes. Positively correlated genes are blue, negatively correlated genes are red; genes are ranked by angular order of eigenvectors. (**J**) STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) analysis of statistically differentially regulated genes in the MEP cluster. Disconnected nodes were removed. (**K**) Pearson correlation matrix of significantly differentially expressed genes present in Pro-B cluster, establishing dependence between multiple genes. Positively correlated genes are blue, negatively correlated genes are red; genes are ranked by angular order of eigenvectors. (**L**) STRING analysis of statistically differentially regulated genes in the Pro-B CD34+ cluster. Disconnected nodes were removed. (**M**) Module score of genes >2 fold-upregulated within AraC-treated CD34+ cells compared to control, all within the CMP cluster. Each dot is one CMP cell. One-way ANOVA. (**N**) Pearson correlation matrix of significantly differentially expressed genes present in CMP cluster, establishing dependence between multiple genes. Positively correlated genes are blue, negatively correlated genes are red; genes are ranked by angular order of eigenvectors. (**O**) STRING analysis of statistically differentially regulated genes in the CMP cluster. Disconnected nodes were removed. (**P**) CMP marker *YBX3* expression visualized using *t*-SNE. (**Q**) *YBX3* expression within the CMP cluster. Each dot is one CMP cell. One-way ANOVA. (**R**) Pan-progenitor proliferation marker *MKI67* (Ki-67) expression visualized using *t*-SNE. Arrows correspond to representative clusters of proliferating cells within Pro-B CD34+ (green), CMP (blue), and MEP (red) clusters.

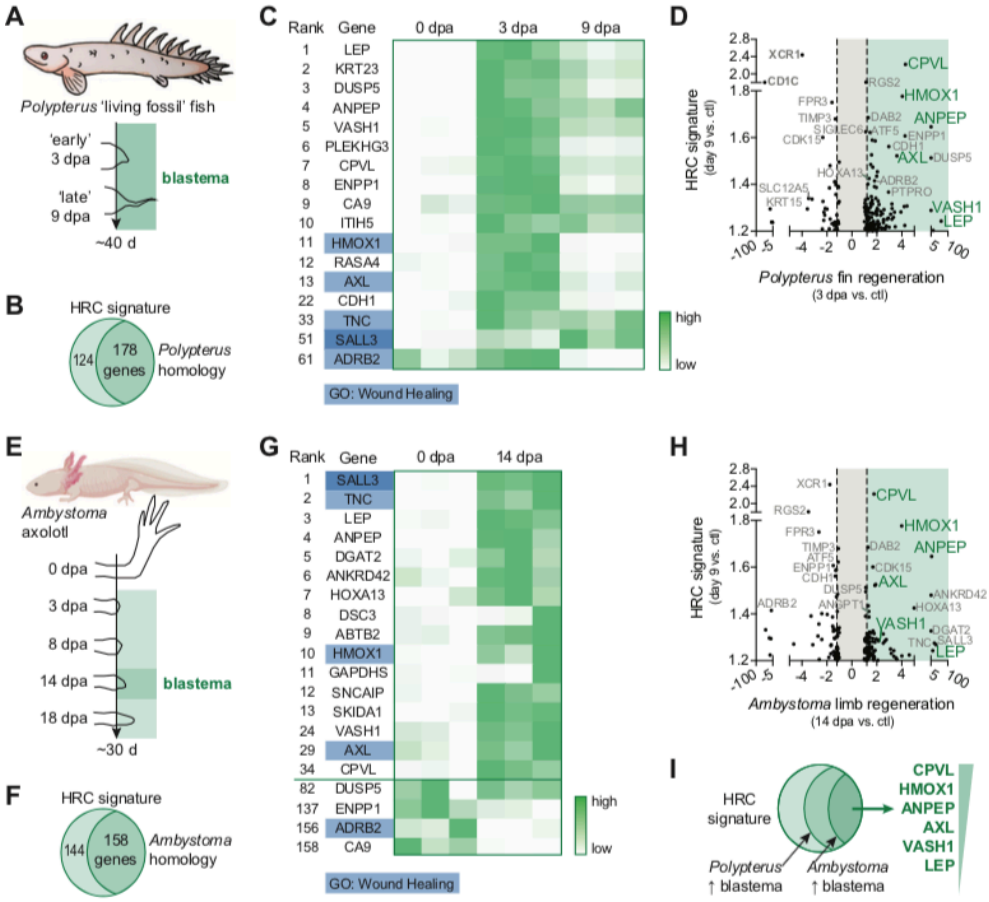


Figure 7. HRC-based evolutionarily conserved early regenerative responses

(**A**) Schematic depicting early and late time points of fin regeneration in *Polypterus* fish. (**B**) Homologous genes shared between *Polypterus* and human HRC signature. (**C**) Identification of up-regulated genes in *Polypterus* fin regeneration which are observed in the HRC signature, ranked by highest differential between 0 and 3 days post-amputation (dpa). (**D**) XY plot showing common up-regulated genes in both *Polypterus* fin and hematopoietic regeneration (green shaded area). Grey shading indicates no change, and white area indicates discordant changes. (**E**) Schematic depicting time points of limb regeneration in axolotl salamanders. Identification of up-regulated genes shared between axolotl limb regeneration and HRC signature. (**F**) Homologous genes shared between axolotl and human HRC signature. (**G**) Identification of up-regulated genes in axolotl blastema which are observed in the HRC signature, ranked by highest differential between 0 and 14 dpa. (**H**) XY plot showing common up-regulated genes in both axolotl blastema and hematopoietic regeneration (green shaded area). Grey shading indicates no change, and white area indicates discordant changes. (**I**) Identification of 6 evolutionarily conserved early response genes to regeneration.

References

Antman, K.S., Griffin, J.D., Elias, A., Socinski, M.A., Ryan, L., Cannistra, S.A., Oette, D., Whitley, M., Frei, E., and Schnipper, L.E. (1988). Effect of Recombinant Human Granulocyte-Macrophage Colony-Stimulating Factor on Chemotherapy-Induced Myelosuppression. New England Journal of Medicine 319, 593-598.

Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature Immunology 20, 163-172.

Aslostovar, L., Boyd, A., Almakadi, M., Collins, T., Leong, D., Tirona, R., Kim, R., Julian, J., Xenocostas, A., Leber, B., et al. (2018). A phase 1 trial evaluating thioridazine in combination with cytarabine in patients with acute myeloid leukemia. Blood Advances 2, 1935-1945.

Bagger, F.O., Kinalis, S., and Rapin, N. (2018). BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. Nucleic Acids Res.

Boyd, A., Aslostovar, L., Reid, J., Ye, W., Tanasijevic, B., Porras, D., Shapovalova, Z., Almakadi, M., Foley, R., Leber, B., et al. (2018). Identification of chemotherapy-induced Leukemic Regenerating Cells reveals a transient vulnerability of human AML recurrence. Cancer Cell 34, 483-498.

Bryant, D.M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M.B., Payzin-Dogru, D., Lee, T.J., Leigh, N.D., Kuo, T.H., Davis, F.G., et al. (2017a). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. Cell Reports 18, 762-776.

Bryant, D.M., Sousounis, K., Payzin-Dogru, D., Bryant, S., Sandoval, A.G.W., Martinez Fernandez, J., Mariano, R., Oshiro, R., Wong, A.Y., Leigh, N.D., et al. (2017b). Identification of regenerative roadblocks via repeat deployment of limb regeneration in axolotls. Regenerative Medicine 2.

Darnet, S., Dragalzew, A.C., Amaral, D.B., Sousa, J.F., Thompson, A.W., Cass, A.N., Lorena, J., Pires, E.S., Costa, C.M., Sousa, M.P., et al. (2019). Deep evolutionary origin of limb and fin regeneration. Proceedings of the National Academy of Sciences of the United States of America.

Echeverri, K., and Tanaka, E.M. (2002). Ectoderm to Mesoderm Lineage Switching During Axolotl Tail Regeneration. Science 298, 1993.

Fares, I., Chagraoui, J., Lehnertz, B., MacRae, T., Mayotte, N., Tomellini, E., Aubert, L., Roux, P.P., and Sauvageau, G. (2017). EPCR expression marks UM171-expanded CD34+ cord blood stem cells. Blood 129, 3344-3351.

Gerber, T., Murawala, P., Knapp, D., Masselink, W., Schuez, M., Hermann, S., Gac-Santel, M., Nowoshilow, S., Kageyama, J., Khattak, S., et al. (2018). Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. Science.

He, S., Roberts, P.J., Sorrentino, J.A., Bisi, J.E., Storrie-White, H., Tiessen, R.G., Makhuli, K.M., Wargin, W.A., Tadema, H., van Hoogdalem, E.J., et al. (2017). Transient CDK4/6 inhibition protects hematopoietic stem cells from chemotherapy-induced exhaustion. Sci Transl Med 9.

Kim, K.S., Zhang, D.-L., Kovtunovych, G., Ghosh, M.C., Ollivierre, H., Eckhaus, M.A., and Rouault, T.A. (2018). Infused wild-type macrophages reside and self-renew in the liver to rescue the hemolysis and anemia of Hmox1-deficient mice. Blood Advances 2, 2732-2743.

Koch, A.E., Polverini, P.J., Kunkel, S.L., Harlow, L.A., DiPietro, L.A., Elner, V.M., Elner, S.G., Strieter, R.M. (1992). Interleukin-8 as a Macrophage-Derived Mediator of Angiogenesis. Science 258, 1798-1801.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research 44, W90-97.

Lu, K., Nakagawa, M.M., Thummar, K., and Rathinam, C.V. (2016). Slicer Endonuclease Argonaute 2 Is a Negative Regulator of Hematopoietic Stem Cell Quiescence. Stem cells 34, 1343-1353.

Lu, Y.C., Sanada, C., Xavier-Ferrucio, J., Wang, L., Zhang, P.X., Grimes, H.L., Venkatasubramanian, M., Chetal, K., Aronow, B., et al. (2018). The Molecular Signature of Megakaryocyte-Erythroid Progenitors Reveals a Role for the Cell Cycle in Fate Specification. Cell reports 25, 2083-2093 e2084.

Meziani, L., Deutsch, E., and Mondini, M. (2018). Macrophages in radiation injury: a new therapeutic target. Oncoimmunology 7, e1494488-e1494488.

Notta, F., Doulatov, S., Laurenti, E., Poeppl, A., Jurisica, I., Dick, J.E. (2011). Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. Science 333, 218-221.

Nowoshilow, S., Schloissnig, S., Fei, J.F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature 554, 50-55.

Parker, K.H., Sinha, P., Horn, L.A., Clements, V.K., Yang, H., Li, J., Tracey, K.J., and Ostrand-Rosenberg, S. (2014). HMGB1 enhances immune suppression by facilitating the differentiation and suppressive activity of myeloid-derived suppressor cells. Cancer Research 74, 5723-5733.

Pellin, D., Loperfido, M., Baricordi, C., Wolock, S.L., Montepeloso, A., Weinberg, O.K., Biffi, A., LKlein, A.M., and Biasco, L. (2019). A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. Nature Communications 10.

Prashad, S.L., Calvanese, V., Yao, C.Y., Kaiser, J., Wang, Y., Sasidharan, R., Crooks, G., Magnusson, M., and Mikkola, H.K. (2015). GPI-80 defines self-renewal ability in hematopoietic stem cells during human development. Cell Stem Cell 16, 80-87.

Reid, J.C., Tanasijevic, B., Golubeva, D., Boyd, A.L., Porras, D.P., Collins, T.J., and Bhatia, M. (2018). CXCL12/CXCR4 Signaling Enhances Human PSC-Derived Hematopoietic Progenitor Function and Overcomes Early In Vivo Transplantation Failure. Stem Cell Reports 10, 1625-1641.

Saxena, M., Balan, S., Roudko, V., and Bhardwaj, N. (2018). Towards superior dendritic-cell vaccines for cancer therapy. Nature Biomedical Engineering 2, 341-346.

Singbrant, S., Mattebo, A., Sigvardsson, M., Strid, T., and Flygare, J. (2020). Prospective isolation of radiation induced erythroid stress progenitors reveals unique transcriptomic and epigenetic signatures enabling increased erythroid output. Haematologica.

Stewart, R., Rascon, C.A., Tian, S., Nie, J., Barry, C., Chu, L.F., Ardalani, H., Wagner, R.J., Probasco, M.D., Bolin, J.M., et al. (2013). Comparative RNA-seq analysis in the unsequenced axolotl: the oncogene burst highlights early gene expression in the blastema. PLoS Comput Biol 9, e1002936.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902.

Tapia, N., Reinhardt, P., Duemmler, A., Wu, G., Arauzo-Bravo, M.J., Esch, D., Greber, B., Cojocaru, V., Rascon, C.A., Tazaki, A., et al. (2012). Reprogramming to pluripotency is an ancient trait of vertebrate Oct4 and Pou2 proteins. Nature Communications 3, 1279.

Theocharidis, G., Bhasin, S.S., Kounas, K., Bhasin, M.K., and Veves, A. (2018). Single Cell RNA-Seq Analyses of Healthy Lower Extremity Skin and Diabetic Foot Ulcers Uncover Distinct Immune Landscape of Diabetic Wound Healing. Diabetes 67, 647-P.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356, eaah4573.

Voss, S.R., Palumbo, A., Nagarajan, R., Gardiner, D.M., Muneoka, K., Stromberg, A.J., and Athippozhy, A.T. (2015). Gene expression during the first 28 days of axolotl limb regeneration I: Experimental design and global analysis of gene expression. Regeneration 2, 120-136.

Xia, C., Braunstein, Z., Toomey, A.C., Zhong, J., and Rao, X. (2018). S100 Proteins As an Important Regulator of Macrophage Inflammation. Frontiers in immunology 8, 1908-1908.

Zappia, L., and Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience 7.

Zheng, G., Terry, J., Belgrader, P., Ryvkin, P., Bent, Z., Wilson, R., Ziraldo, S., Wheeler, T., McDermott, G., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications 8, 1-12.

Supplemental Figures

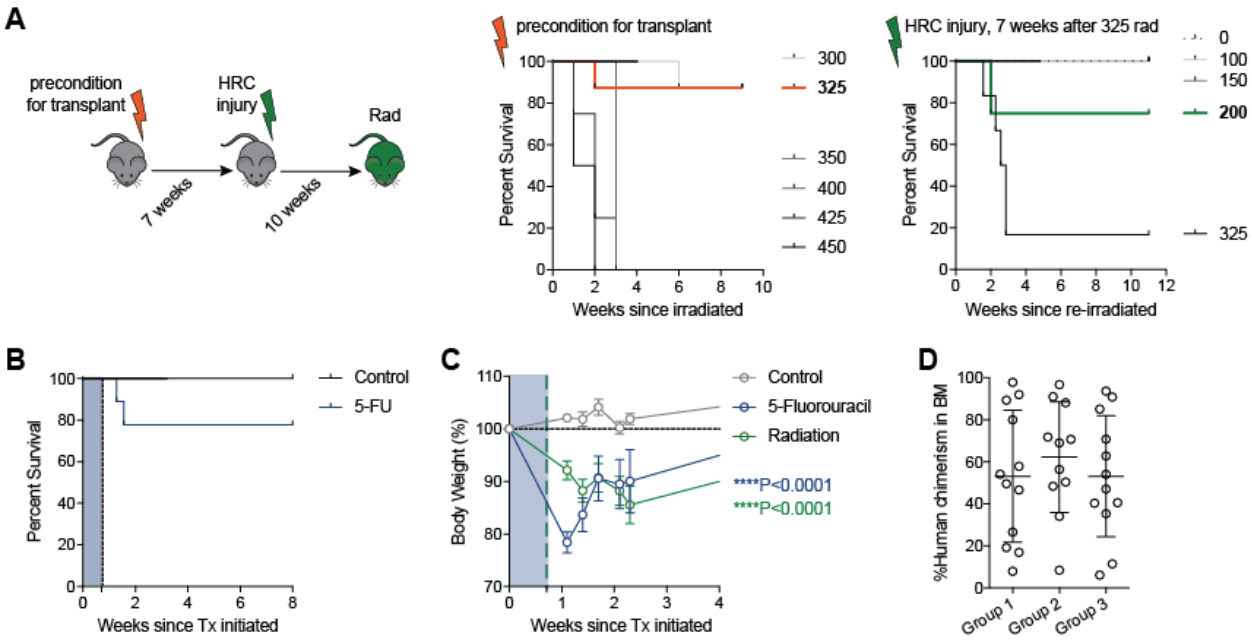


Figure S1. Optimized maximal tolerable dose of myelosuppressive treatments

(**A**) All mice were radiated the day before transplantation with human HSCs, this dose was determined to be 325 rad for NOD/SCID mice. However, this dose was lethal when administered again 7 weeks later. Therefore, 200 rad was determined to be the maximal tolerable dose which induced 20% mortality. (**B**) Five daily consecutive intraperitoneal (IP) injections of 30 mg/kg 5-FU were determined to be the maximal tolerable dose which induced 20% mortality. (**C**) Transient weight loss is observed for both myelosuppressive treatments, 5-FU and radiation. (**D**) Representative experiment which randomized treatment mice by engraftment level using human chimerism of BM aspirate taken 7 wk post-transplant.

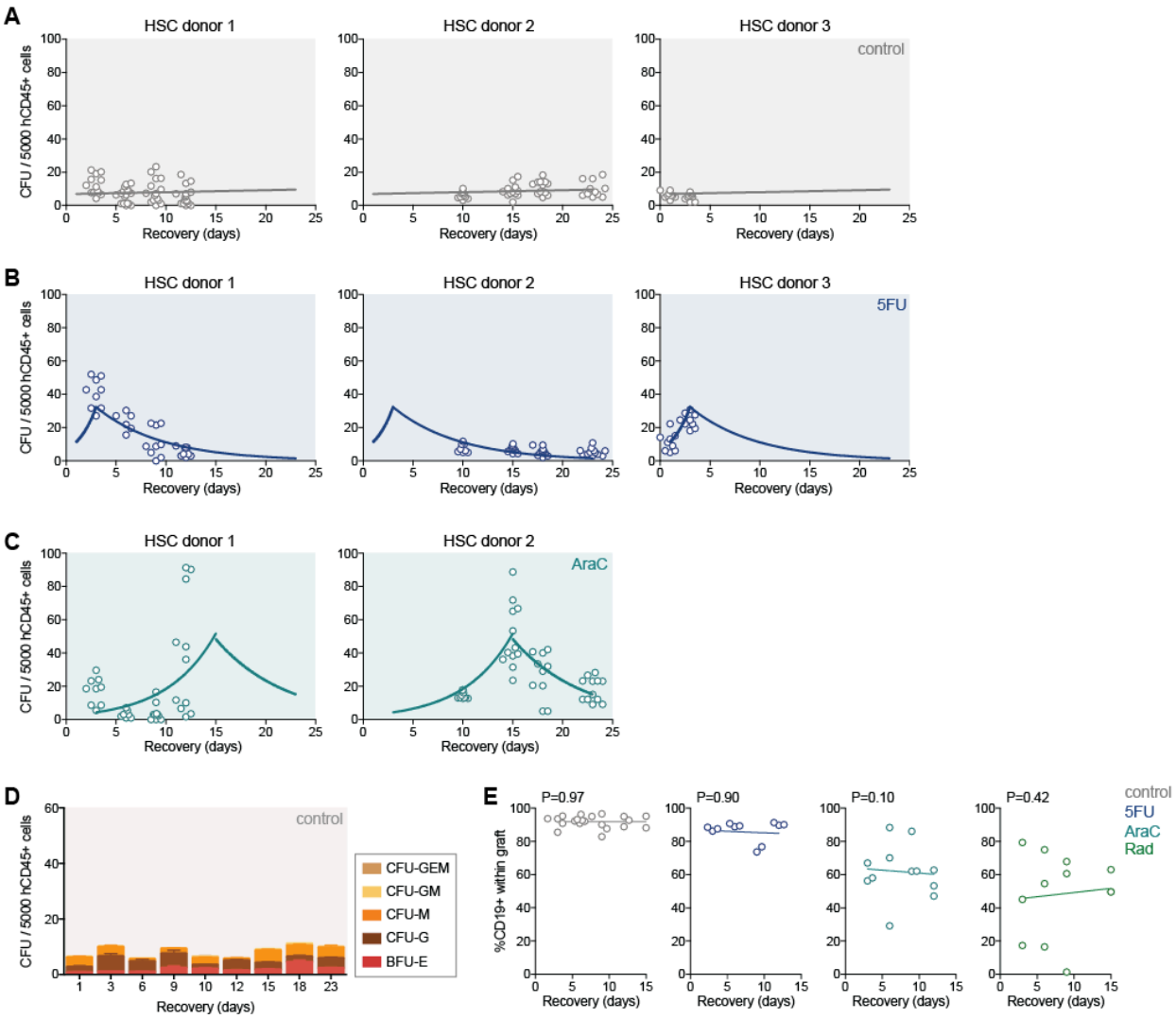


Figure S2. Independent HSC donors corroborate temporal regulation of HRPC

(**A**) Three HSC donors were transplanted in 3 independent experiments, indicated with 3 plots. Total CFU counted per 5000 human cells seeded in methylcellulose, of human cells purified from individual control-Tx xenografted mice harvested at the indicated time points. Combined data is shown in Figure 2H. (**B**) Three HSC donors were transplanted in 3 independent experiments, indicated with 3 plots. Total CFU counted per 5000 human cells purified from individual 5-FU-Tx xenografted mice harvested at the indicated time points. Combined data is shown in Figure 2I. (**C**) Three HSC donors were transplanted in 3 independent experiments, indicated with 3 plots. Total CFU counted per 5000 human cells purified from individual AraC-Tx xenografted mice harvested at the indicated time points. Combined data is shown in Figure 2J. (**D**) CFU lineage composition of human cells purified from individual control-Tx xenografted mice. (**E**) CD19+ composition within the human graft of individual mice harvested at the indicated time points. Each dot is one mouse. Linear regression determined slope did not differ significantly from zero.

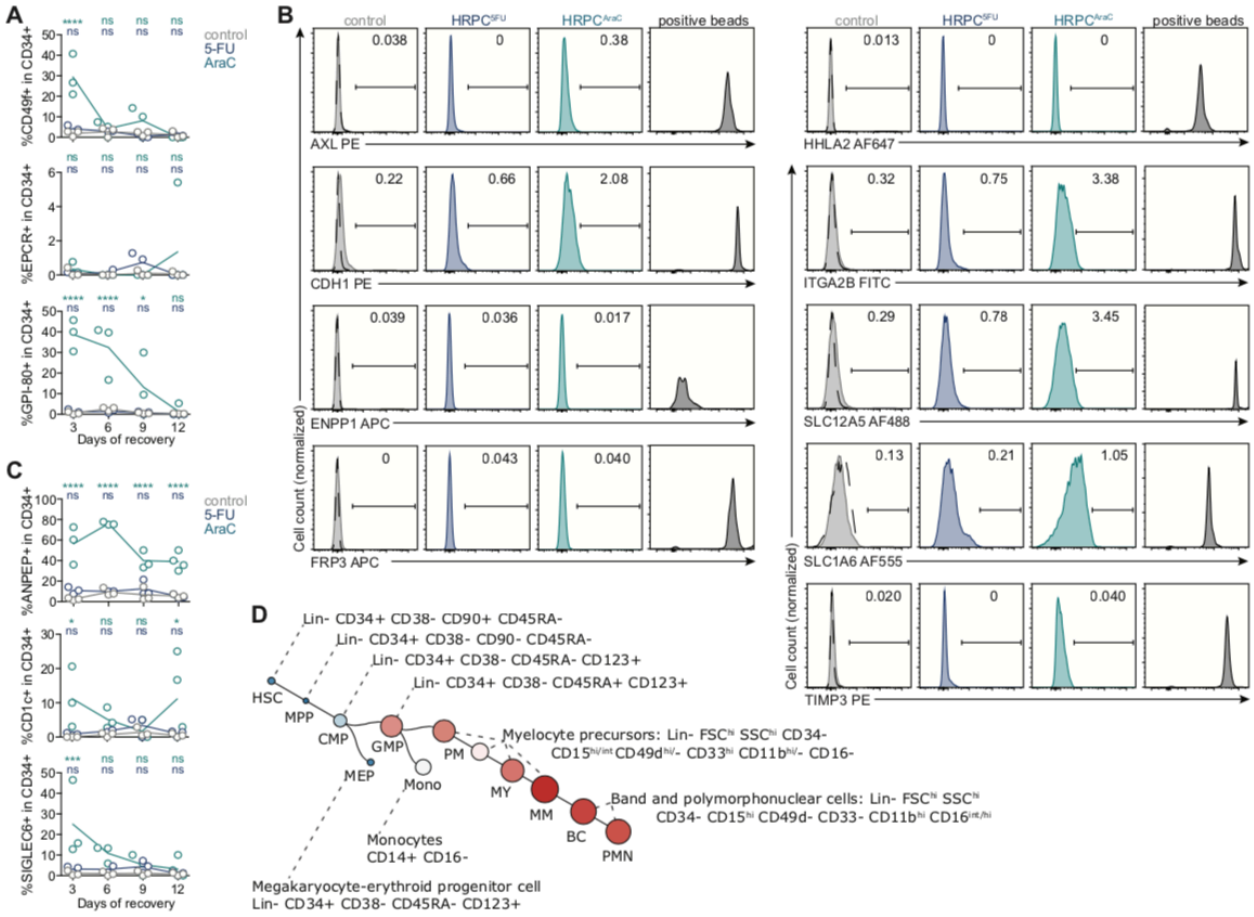


Figure S3. Expanded analysis of candidate markers of HRPCs

(**A**) Quantitation of classical HSC markers within CD34+CD45+ primitive population of each xenograft. Each dot is one mouse. Two-way ANOVA vs. control Tx. (**B**) Evaluation of candidate HRC cell surface markers on whole BM from xenografted mice. Positive beads were used to demonstrate effective antibody binding and detection of fluorescence. Each dot is one mouse. Two-way ANOVA vs. control Tx. (**C**) Quantitation of candidate HRC markers within CD34+CD45+ primitive population of each xenograft. Each dot is one mouse. Two-way ANOVA vs. control Tx. (**D**) Expanded view of hierarchical differentiation tree (phenotypes and abbreviations; Bloodspot.com) representing discrete hematopoietic subpopulations. For all panels, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

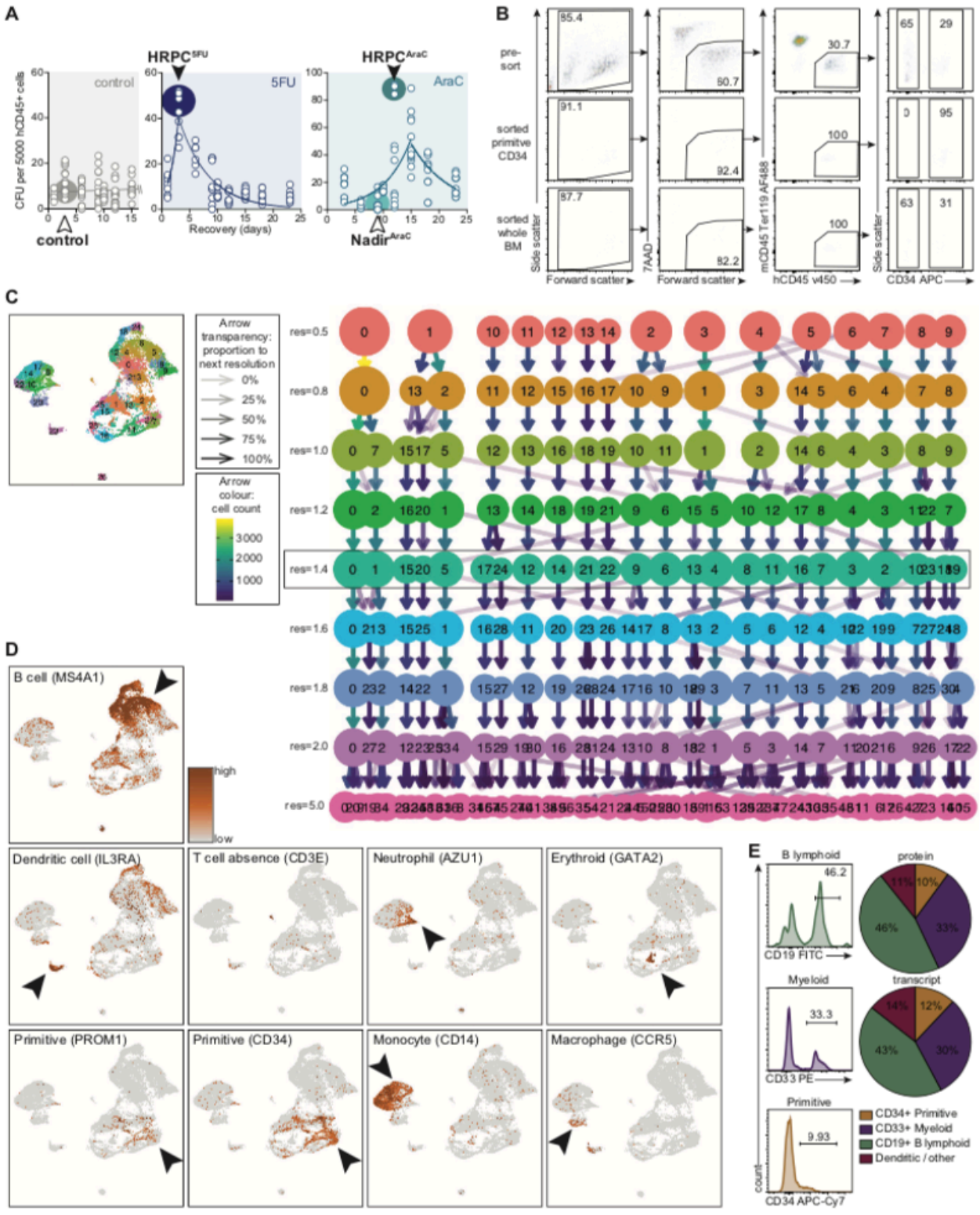


Figure legend on next page

Figure S4. Expanded scRNA-seq quality control analysis of whole BM samples

(**A**) Representation of the CFU data shown in Figure 2H-J, where the specific xenograft sample(s) used for scRNA-seq are indicated with dark shading. All samples originated from the same HSC donor. (**B**) Representative FACS purification of cells from mouse BM which were then immediately loaded in the Chromium Controller (10X Genomics). (**C**) After QC, 24,616 human whole BM cells were visualized using UMAP (*left*), the appropriate resolution was visualized with *clustree* and determined to be 1.6 (*right*). (**D**) Expression plots (UMAP) showing key transcriptional markers of the main hematopoietic cell identities. Arrows indicate the main subpopulation expressing each marker. Scale is based on UMI counts. (**E**) Flow cytometric plots of xenograft cells gated within hCD45+, from $HRPC^{AraC}$ time point, on the day of harvest, and represented as a pie chart (protein) comprising the total graft. Additionally, this same sample was prepared for scRNA-seq and the subpopulations ascribed to each of the main hematopoietic lineages (Figure 4B) is shown summarized here for the $HRPC^{AraC}$ sample.

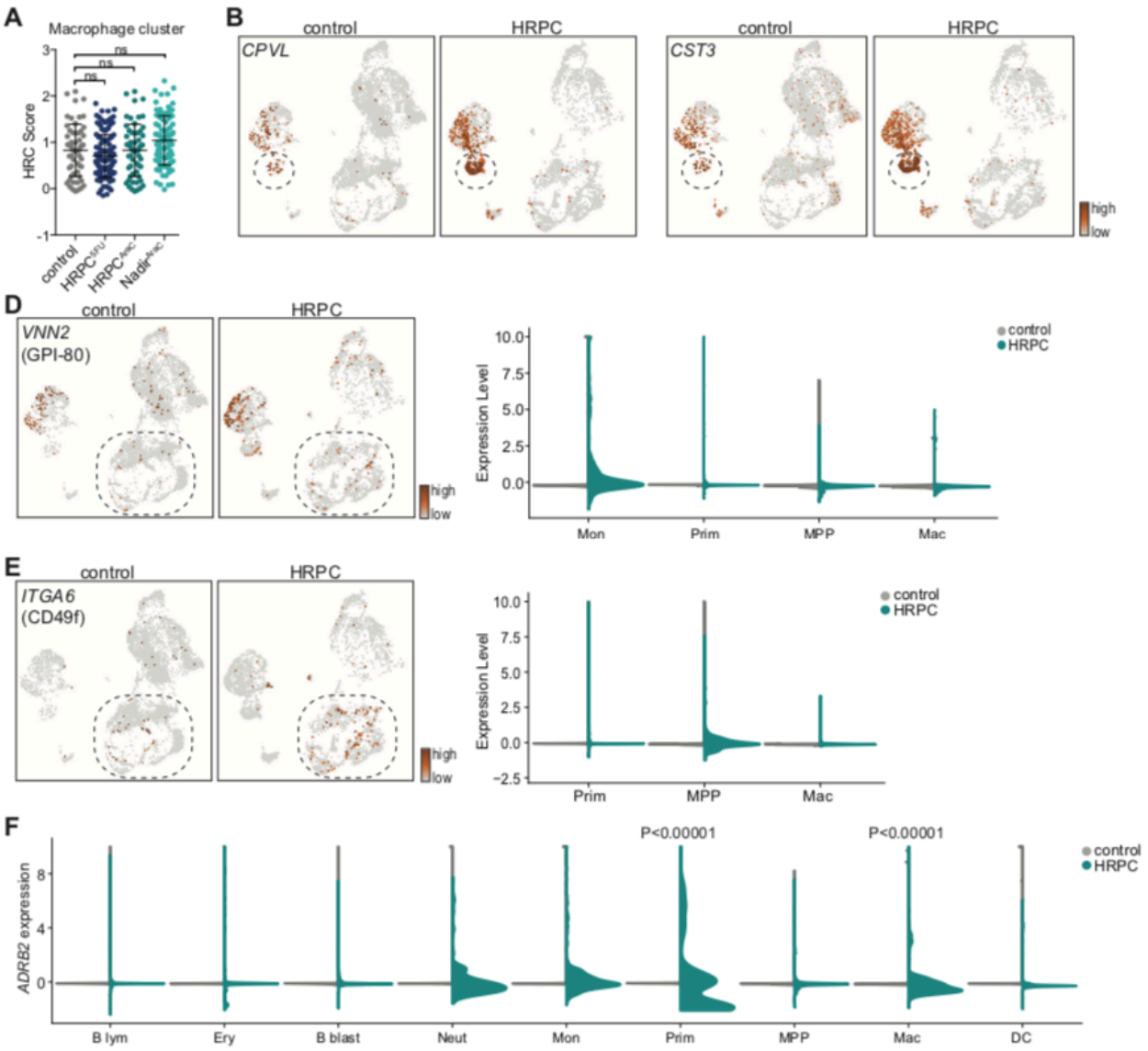


Figure S5. Expanded scRNA-seq analysis of whole BM markers of HSC and HRCs

(**A**) Module score of the HRC signature genes (Table S4) for each cell within the macrophage cluster of each of the four whole BM samples. Kruskal-Wallis one-way ANOVA. (**B**) Macrophage-associated HRC gene expression visualized using UMAP in R. HRPC refers to $HRPC^{AraC}$ whole BM sample. (**D,E**) Classical HSC marker expression of whole BM cells visualized using UMAP in R, and violin plots of cell identity clusters with positive cells. HRPC refers to $HRPC^{AraC}$. (**F**) Violin plot of *ADRB2* expression per cell identity cluster. HRPC refers to $HRPC^{AraC}$. Fisher's exact test.

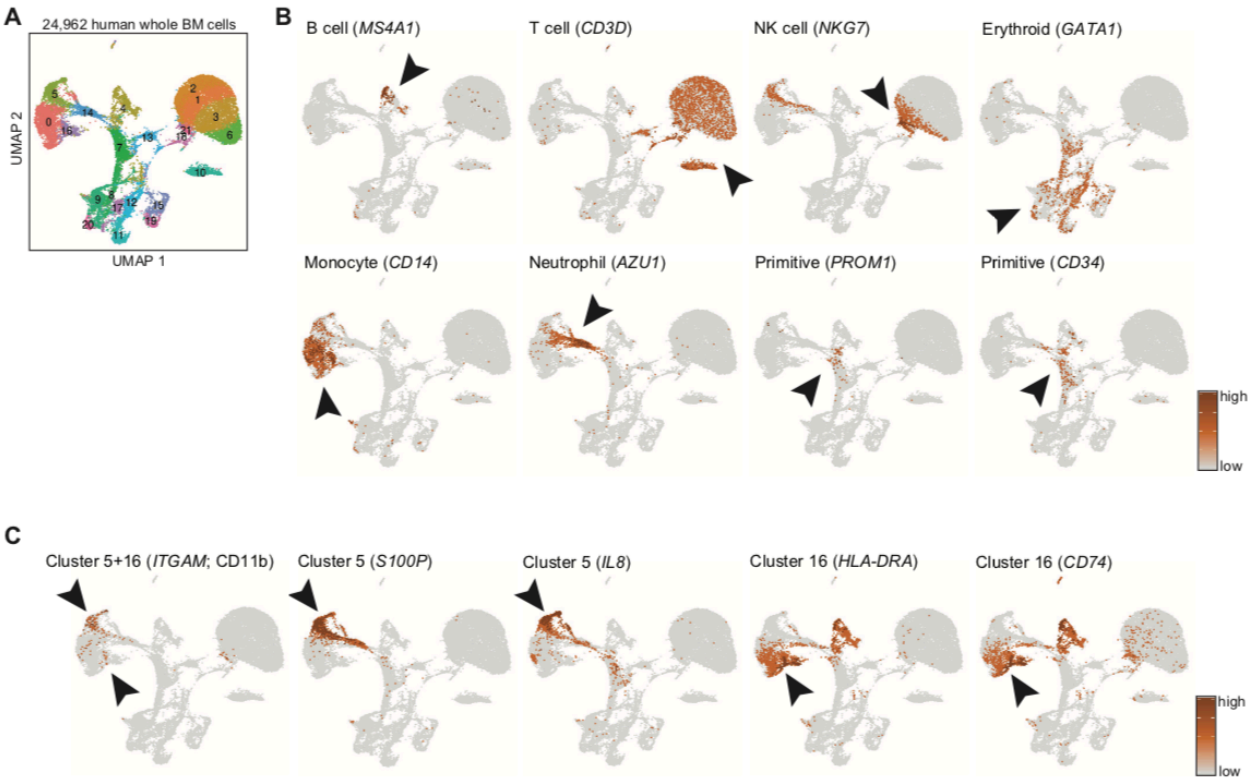


Figure S6. Expanded scRNA-seq analysis of whole BM samples from AML patients

(**A**) Clustering of whole BM cells visualized using UMAP in R. (**B**) Expression plots (UMAP) showing key transcriptional markers of the main hematopoietic cell identities. Arrows indicate the main subpopulation expressing each marker. Scale is based on UMI counts. (**C**) Expression plots (UMAP) showing co-expression of CD11b (*ITGAM*) as well as significantly differentially expressed genes between classical macrophages and *S100P*+ 'stressed' macrophages. Arrows indicate the main subpopulation expressing each marker. Scale is based on UMI counts.

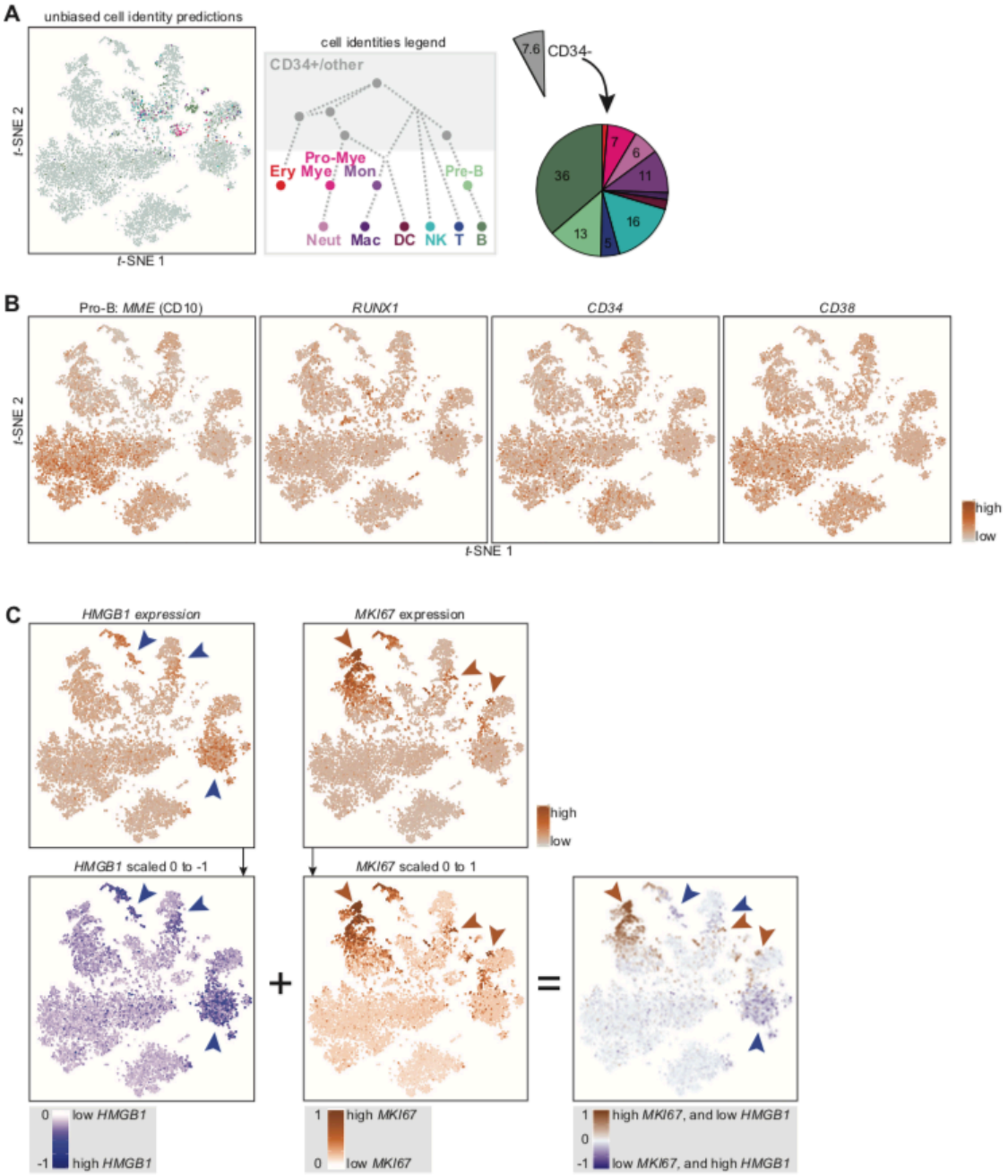


Figure S7. Expanded scRNA-seq analysis of CD34-sorted HRPCs

(**A**) *SingleR* based cell identity predictions of the CD34- cells, and their relative abundance, from CD34+ sorted cells purified from HRPC xenografts. (**B**) Gene expression visualization of progenitors displaying nuanced differences in gene expression, requiring sophisticated cell identity predictions beyond classical manual approaches. (**C**) Gene expression visualization of MKI67 and *HMGB1* and the process to scale and map these two genes, ultimately presented as a single score to highlight diametrically opposed expression patterns.

Supplemental Tables

Table S1. List of candidate HRC markers with flow cytometry-supported antibodies

HRC Rank	Fold-Change	P-value	Transcript	Alternative name(s)	Gene Ontology
1	2.438	0.035	XCR1	GPR5, Lymphotactin Receptor, CCXCR1	2nd in GO_RECEPTOR_ACTIVITY
6	1.804	0.034	CD1C	Differentiation Antigen CD1-Alpha-3, T-Cell Surface Glycoprotein CD1c, Blood Dendritic Cell Antigen 1	
8	1.750	0.041	FPR3	Formyl Peptide Receptor 3	4th in GO_RECEPTOR_ACTIVITY
10	1.680	0.010	TIMP3	MIG-5 Protein	1st in GO_WOUND_HEALING
12	1.646	0.006	ANPEP	CD13, Aminopeptidase M	13th in GO_RECEPTOR_ACTIVITY
14	1.625	0.008	SIGLEC6	Sialic Acid Binding Ig Like Lectin 6; Obesity-Binding Protein 1; CD33 Antigen-Like 1 (CD33L), CD327	1st in GO_CELL_CELL_SIGNALING
15	1.607	0.048	ENPP1	ARHR2 (Autosomal recessive hypophosphatemic rickets type 2)	5th in GO_RECEPTOR_ACTIVITY
20	1.562	0.014	CDH1	Epithelial Cadherin, E-Cad	
23	1.522	0.050	AXL	AXL receptor tyrosine kinase	4th in GO_WOUND_HEALING; 11th in GO_RECEPTOR_ACTIVITY
37	1.441	0.015	ITGA2B	Integrin Subunit Alpha 2b, CD41, Platelet Fibrinogen Receptor Alpha Subunit	6th in GO_WOUND_HEALING
42	1.417	0.013	SLC1A6	Solute Carrier Family 1 (High Affinity Aspartate/Glutamate Transporter) Member 6, EAAT4, Excitatory Amino Acid Transporter 4	2nd in GO_CELL_CELL_SIGNALING
44	1.415	0.037	ADRB2	Beta-2 Adrenoreceptor, B2AR, BAR	7th in GO_WOUND_HEALING; 18th in GO_RECEPTOR_ACTIVITY
65	1.348	0.015	HHLA2	Human Endogenous Retrovirus-H Long Terminal Repeat-Associating Protein 2, B7y, B7H7	
67	1.340	0.004	SLC12A5	KCC2, Erythroid K-Cl Cotransporter 2, Neuronal K-Cl Cotransporter, EIG14, HKCC2	6th in GO_CELL_CELL_SIGNALING

Table S2. QC metrics of single cell RNA-seq data sets produced from xenografts

Sample ID	cell viability	Cell Ranger		Seurat		
		reads/cell	genes/cell	total samples	samples passing QC	min.cells setting
control whole BM	98	17,926	642	6850	6640	6
HRPC^{5FU} whole BM	95	12,627	772	8649	8649	6
HRPC^{AraC} whole BM	93	20,018	1,582	4765	4273	4
Nadir^{AraC} whole BM	97	18,020	1,172	5474	5055	5
control CD34	96	23,110	943	5386	5174	5
HRPC^{5FU} CD34	94	35,925	1,025	3132	2991	3
HRPC^{AraC} CD34	97	11,736	91	595	589	1
Nadir^{AraC} CD34	84	79,877	2,424	1334	880	1

Table S3. Ranked Enrichr statistics of each scRNA-seq integrated whole BM cluster

Top significant conserved genes of each cluster were assessed using Enrichr Human Gene Atlas database (https://amp.pharm.mssm.edu/Enrichr/), and significant entries were included in this table. Column 1 is colour-coded to the cell identity used in Figure 4C (CD34 uncommitted, brown; CD34 lymphoid & erythroid, goldenrod; Monocyte, violet; Neutrophil, light pink; Macrophage, dark violet; Dendritic cell (A), pink; Dendritic cell (B), dark pink; B lymphoblast, light olive; B lymphocyte, dark olive; Unassigned/high ribosomal gene content, grey). No significant entries returned for cluster 26 and 27; first three entries are shown (P>0.05, not significant).

Cluster	Name	P-value	Score
0	CD19+ B Cells (neg. sel.)	0.00005222	174.05
	Lymphoma burkitts (Daudi)	0.0000524	387.06
	Lymph node	0.02017	191.83
	Lymphoma burkitts (Raji)	0.01554	42.65
1	BDCA4+ Dendritic Cells	0.001821	35.40
	CD34+	0.00566	22.28
2	Thymus	0.0005036	446.69
	CD34+	0.003381	52.93
	Leukemia lymphoblastic (MOLT-4)	0.05747	48.42
3	CD19+ B Cells (neg. sel.)	0.00002877	101.52
	Lymphoma burkitts (Daudi)	0.001162	97.23
	Caudate nucleus	0.09855	22.39
	BDCA4+ Dendritic Cells	0.03732	13.29
4	CD19+ B Cells (neg. sel.)	3.71E-14	692.89
	Tonsil	0.000004971	1083.9
	Lymph node	0.001052	285.1
	Lymphoma burkitts (Raji)	0.01065	29.52
5	CD19+ B Cells (neg. sel.)	1.81E-12	463.21
	Lymphoma burkitts(Raji)	0.00002649	104.78
	Lymphoma burkitts(Daudi)	0.001679	81.12
	Lymph node	0.06107	44.45
6	CD19+ B Cells (neg. sel.)	8.79E-20	1135.99
	Tonsil	0.000007735	905.36
	Lymph node	0.001402	236.75
	Lymphoma burkitts (Raji)	0.0001694	81.53
	Lymphoma burkitts (Daudi)	0.01837	38.34
	Thymus	0.09719	22.85
7	Thymus	3.41E-07	509.32
	CD34+	8.32E-07	100.95
	Leukemia lymphoblastic (MOLT-4)	0.00208	73.02
	Caudate nucleus	0.009669	62.54
	CD19+ B Cells (neg. sel.)	0.01163	20.11
	Lung	0.02629	16.98
8	CD33+ Myeloid	1.01E-12	443.79
	CD14+ Monocytes	0.00005051	117.42
9	CD19+ B Cells (neg. sel.)	0.00398	50.3
	Leukemia chronic Myelogenous K-562	0.004323	109.76
	Cingulate Cortex	0.02609	138.12
10	CD33+ Myeloid	1.88E-16	639.95
	CD14+ Monocytes	1.49E-11	520.68
	Whole Blood	2.66E-10	343.13
	Bone marrow	0.001914	192.57
11	721 B lymphoblasts	0.000003507	40.7
	Lymphoma burkitts (Raji)	0.00006004	57.65

Cluster	Name	P-value	Score
	CD34+	0.0001592	35.68
	Lymphoma burkitts (Daudi)	0.0001876	81.23
12	CD19+ B Cells (neg. sel.)	4.71E-11	546.76
	Tonsil	8.58E-09	3007.93
	Lymph node	0.000005533	1033.12
	Lymphoma burkitts (Raji)	0.04404	18.52
13	Fetal Thyroid	0.003064	140.32
	Fetal brain	0.007503	36.13
	CD34+	0.04535	11.51
14	CD33+ Myeloid	2.37E-10	391.69
	CD14+ Monocytes	2.05E-07	321.66
	Whole Blood	0.00002667	136.6
15	CD71+ Early Erythroid	7.38E-08	215.97
	721 B lymphoblasts	0.000001981	77.37
	Leukemia lymphoblastic (MOLT-4)	0.000007758	362.61
	CD105+ Endothelial	0.00745	36.11
	CD34+	0.03259	14.48
16	721 B lymphoblasts	6.42E-11	144.1
	CD71+ Early Erythroid	2.49E-10	252.54
	Leukemia lymphoblastic (MOLT-4)	0.00000282	284.99
	CD105+ Endothelial	0.00006443	82.36
	CD34+	0.03326	11.11
17	CD33+ Myeloid	7.02E-20	806.31
	Whole Blood	0.00008036	75.91
	CD14+ Monocytes	0.000204	76.51
18	CD19+ B Cells (neg. sel.)	1.26E-16	661.34
	Tonsil	0.00002325	572.59
	Lymphoma burkitts (Raji)	0.0009476	45.6
	Lymph node	0.002867	147.19
	Lymphoma burkitts (Daudi)	0.03597	22.25
19	CD19+ B Cells (neg. sel.)	0.0001369	86.37
	Lymphoma burkitts (Daudi)	0.0006755	126.05
	Leukemia chronic Myelogenous K-562	0.01042	58.89
20	CD14+ Monocytes	1.04E-14	588.45
	BDCA4+ Dendritic Cells	3.43E-08	173.62
	CD33+ Myeloid	6.00E-08	134.7
	Whole Blood	0.0005112	44.23
	CD19+ B Cells (neg. sel.)	0.009035	22.85
	Olfactory Bulb	0.01198	53.3
	Lung	0.02175	19.21
21	CD33+ Myeloid	3.04E-11	222.91
	Lymphoma burkitts (Raji)	0.01028	21.49
	Thymus	0.01162	54.6
	CD19+ B Cells (neg. sel.)	0.01693	16.5
	Whole Blood	0.03448	10.92
22	CD33+ Myeloid	7.57E-24	1120.08
	CD14+ Monocytes	1.10E-15	835.11
	Whole Blood	4.20E-11	365.24
	Bone marrow	0.0024	165.72
23	BDCA4+ Dendritic Cells	9.37E-17	835.05
	CD56+ NK Cells	0.01337	18.36
24	CD19+ B Cells (neg. sel.)	9.34E-10	363.34
	Lymphoma burkitts (Raji)	0.000003992	168.08
	Lymph node	0.0009726	299.92
	BDCA4+ Dendritic Cells	0.003103	37.34
	Lymphoma burkitts (Daudi)	0.01296	50.03
25	721 B lymphoblasts	0.0001285	58.07
	CD71+ Early Erythroid	0.0002398	100.49

Cluster	Name	P-value	Score
	CD34+	0.005908	39.78
26	Thyroid	0.1617	10.44
	Heart	0.172	9.43
	CD71+ Early Erythroid	0.2231	6.03
27	Leukemia chronic Myelogenous K-562	0.05445	52.16
	Uterus	0.08691	27.01
	Prostate	0.1066	19.98
28	721 B lymphoblasts	1.09E-16	288.33
	CD71+ Early Erythroid	3.01E-14	444.48
	CD34+	0.03326	11.11

Table S4. HRC signature genes detected in scRNA-seq whole BM cells

HRC genes					
ABTB2	CPVL	GPRC5C	MITF	PTPN1	SNX3
ACSS2	CST3	GPT	MMP16	PTPN6	TAF13
ALKBH2	CTTNBP2NL	GTSF1L	MPEG1	RAB27B	TANC1
ANPEP	CXCL16	HLA-DQB2	MTRNR2L10	RAB30	TAS1R3
AOC1	CYSTM1	HLA-DRA	MYO1D	RASA4	TBC1D9
ARMCX2	DAB2	HLA-DRB1	MYRF	RBPMS	TIMP3
ATF5	DAPP1	HMOX1	NCR3	RELB	TNC
ATP2B1	DCAF4	HTR7	NEO1	RGS2	TREM1
ATP6AP1L	DLC1	IFNGR2	NIPAL2	ROPN1L	TRPV3
C1orf162	ENPP1	IL27	NRIP3	RUNDC3B	TSPAN10
C1orf54	FAH	IL4R	NRP2	SARAF	WWC2
CA9	FANK1	IL5RA	NRSN2	SIGLEC6	XCR1
CAMSAP2	FAT4	KCND3	NUDT2	SKIDA1	ZBTB47
CD1A	FBXO16	KRT23	NUDT8	SLC22A18	ZBTB8B
CD1C	FERMT2	LIF	OGN	SLC25A53	ZNF300
CDH23	FPR3	LMNA	PDE6B	SLC45A3	ZNF503
CDK15	FSTL3	MAP1LC3B2	PDGFC	SMOX	
CEP152	GLIS3	MAP3K8	PDK4	SMPDL3B	
CFH	GOLGA8M	MARVELD2	PGF	SNAP23	
CLNK	GOLM1	MCC	PKIB	SNTB1	

Table S5. SingleR predicted cell identities of CD34+ sorted scRNA-seq data

	CD34+ sorted scRNA-seq sample				visualization schema		
SingleR Identity	**control**	**5-FU d3**	**AraC d12**	**AraC d9**	**CD34+**	**CD34-**	**within CD34**
Astrocyte	1	1	4	0	azure2	azure2	
B cell	67	102	41	36	azure3	darkseagreen4	
Bone Marrow	6	3	4	0	azure3	azure3	
BM & Prog.	1	0	8	0	azure3	azure3	
Chondrocytes	0	1	2	0	azure3	azure3	
CMP	69	42	144	570	darkgoldenrod	azure2	deepskyblue4
Dendritic Cell	0	0	14	2	azure3	deeppink4	
Embryonic stem cells	2	0	2	0	azure2	azure2	
Endothelial cells	0	0	3	0	azure3	azure3	
Epithelial cells	0	0	3	1	azure3	azure3	
Erythroblast	1	0	7	1	azure3	darkred	
Fibroblasts	1	1	3	0	azure2	azure2	
GMP	25	7	71	87	darkgoldenrod	azure2	darkturquoise
HSC -G-CSF	0	1	0	1	darkorange4	azure2	darkorange3
HSC CD34+	64	45	25	11	darkorange4	azure2	darkorange4
Keratinocytes	1	0	3	0	azure2	azure2	
Macrophage	0	0	8	4	azure3	darkorchid4	
MEP	3	4	28	6	darkgoldenrod	azure2	red
Monocyte	7	24	10	31	azure3	darkorchid1	
MSC	0	0	1	0	azure3	azure3	
Myelocyte	1	0	1	0	azure3	deeppink2	
Neurons	1	1	6	1	azure2	azure2	
Neutrophil	7	12	0	0	azure3	violet	
Neutrophils	3	15	1	5	azure3	violet	
NK cell	24	36	18	31	azure3	cyan3	
Osteoblasts	0	0	1	0	azure3	azure3	
Pre-B cell CD34-	22	32	28	9	azure3	darkseagreen3	
Pro-B cell CD34+	4847	2645	112	70	darkgoldenrod	azure2	darkseagreen3
Pro-Myelocyte	12	13	12	10	azure3	deeppink	
Smooth muscle cells	4	2	12	0	azure3	azure3	
T cells	6	4	17	4	azure3	blue	
Total number of cells	**5175**	**2991**	**589**	**880**			

	highest
	middle
	lowest

Chapter 5: Discussion

5.1 Overview of thesis

HSC transplantation has provided many benefits for patients as well as opening new fields of research; hematopoietic hierarchy, regeneration, and other stem cell applications (Talib and Shepard, 2020). There remains, however, questions of how to meet the shortages of HSCs for donation as well as addressing endogenous dysfunction (i.e. CIM for non-hematopoietic reasons, or CLL manifestation) where HSC transplantation is not a viable option. HSC biological research is strongly based on functionally described cells (Becker et al., 1963), with beautiful work describing basic and complex biology in 'patient surrogates'; immuno-compromised mice which can support continued human hematopoiesis (Bhatia et al., 1997). Efforts to expand HSCs have mainly focussed on propagating CB HSCs *in vitro* (Fares et al., 2017), or attempting to propagate hPSC-HPC *in vivo* (Doulatov et al., 2013), where neither is in its preferred habitat. The success and limitations of both modalities is functionally defined in xenografts prior to any clinical consideration. *De novo* patient samples of large patient cohorts are also extremely valuable, but without the experimental means to reduce noise and evaluate causal roles of interventions, these large data sets are correlative at best and may be inaccurate; for example, when up to 15% of the cells of each sample are not the cells of interest, then noise is introduced downstream when bulk RNA-seq and epigenetic analyses are performed. This presents a missed opportunity as the unquestionable value of *de novo* samples cannot be leveraged to their full potential. Unfortunately, this is a common occurrence and often unavoidable since prospective isolation is not always available.

This thesis offers experimental evidence of new models to functionally describe and target transitions in the hematopoietic hierarchy; by identifying and targeting gene products arising from leukemic mutations (***Chapter 2***), by identifying and restoring CXCR4 signaling to hPSC-HPC to improve their functional capacity (***Chapter 3***), and identifying and functionally describing hematopoietic regenerating progenitor cell (HRPC) biology in xenografts combined with single cell RNA-seq (***Chapter 4***). Together, these findings support the hypothesis that human stem cell models provide a focused approach to interrogate the regulation of hematopoiesis from the apex of the hierarchy, which can be used to understand the promotion of healthy hematopoiesis and understand malignant transformation. By parsing apart these complex issues, my work has identified actionable targets and pathways to advance our understanding of hematopoiesis *in situ*.

5.2 Common themes between PSCs and cancer

Despite the wealth of knowledge from the study of human cancer 'omics', identifying selective targetable features of individual cancers remains a major challenge (Lin et al., 2019). Previously, our lab has described and pioneered the use of a variant line of H9 human ESCs with features of cancerous transformation (termed H9V1), since they exhibit enhanced pluripotent capacity and reduced differentiation capacity (Werbowetski-Ogilvie et al., 2009). H9V1 cells have since been used to enrich for compounds which target transformed leukemic progenitors but do not significantly affect healthy hematopoietic progenitors (Sachlos et al., 2012), with enough supporting evidence to permit a phase 1 clinical trial (Aslostovar et al., 2018). However, no gross chromosomal aberrations were reported in H9V1 (Werbowetski-

Ogilvie et al., 2009), suggesting PSCs and cancer may share pathways not directly linked by the same mutations. Secondly, *in silico* analysis of human PSCs and thousands of primary cancer tumors have lead researchers to postulate shared signalling pathways exist between PSCs and cancerous tumors (Malta et al., 2018). Through the repurposing of a machine learning technique to identify mechanistically linked genes (Sokolov et al., 2016a; Sokolov et al., 2016b), Malta *et al*., reported core pathways which could be used to rank cancer tumors progressively from 'more to less differentiated', with the 'less differentiated' cancers associated with a worse prognosis (Malta et al., 2018). Third, capture of AML-related leukemic mutations in human induced PSCs showed that fundamental aspects of hematopoietic differentiation may be reflected (Kotini et al., 2017), but the biology of these mutations was not investigated while in the pluripotent state. These three lines of thought lead me to investigate whether *leukemic mutations captured in the pluripotent state reflect clinically relevant malignant regulation of hematopoiesis.*

My work in ***Chapter 2*** biologically links patient CLL and human PSCs via a specific and commonly shared genetic lesion; tri12. I first validated this assertion using the same machine learning approach originally developed by Sokolov *et al*. (Malta et al., 2018; Sokolov et al., 2016a; Sokolov et al., 2016b) where I developed a tri12 score based on differences between tri12 and isogenic PSCs. I then used this score to classify CLL patients based on the gene expression datasets of highly purified CLL cells (Morabito et al., 2013), generously shared by my new Italian collaborators and applied in Chapter 2. As opposed to generalized claims of 'embryonic dedifferentiation' in pan-cancer analyses, this approach shows that the reduction of 'background noise' through the use of an isogenic-controlled PSC provided a short list of specific actionable targets derived from increased gene dosage induced by tri12, which were translatable to primary CLL at a high success rate. Therefore, the use of human PSCs harboring a complex leukemic mutation is a suitable and valuable model for cleanly dissecting and investigating the pathobiological properties of a leukemic mutation.

5.2.1 Future clinical perspectives: modelling malignant hematopoiesis

I began this work largely serendipitously as an early graduate student learning to culture and differentiate multiple hPSC lines (ESC and iPSC) from our lab's cryopreserved bank. Reproducibly, a very high passage (>150 weeks in culture) CA2 human ESC line produced unprecedented frequencies of hematopoietic differentiation, whereas the low passage (<50 weeks in culture) CA2 lines were within normal hematopoietic differentiation ranges similar to H9 ESCs and CB iPSCs. Multiple genetic tests identified the sole acquisition of tri12 in the 'high passage CA2, which were then reclassified as tri12 PSC for all subsequent work. Around the same time, beautiful work by Banno *et al*., demonstrated that iPSC derived from Down's syndrome patients (trisomy 21) could be used to study aberrant hematopoietic differentiation *in vitro* (Banno et al., 2016). Since Down's syndrome patients may experience transient myeloproliferative disorder, the authors attributed increased gene dosage of a few key genes on chromosome 21 to alterations in hematopoietic differentiation. Secondly, spontaneous acquisition of tri12 is strongly associated in the medical literature with CLL, and researchers have even remarked how incredibly restricted trisomy 12 is to CLL (Mitelman and Levan, 1981). Spontaneous acquisition of tri12 is also quite common for human PSCs (De Sousa et al., 2016; Draper et al., 2003; Gertow et al., 2007); potentially a useful coincidence.

Together these data lead me to *hypothesize that tri12 captured in the pluripotent state reflects clinically relevant malignant regulation of CLL*.

Tri12 represents a transcriptionally distinct subpopulation of CLL patients (Beekman et al., 2018). However, these and other studies have not fully purified the leukemic cells from bulk patient samples to achieve certainty in the targets identified, and drug screening campaigns have resulted in perceived 'patient-specific' responses instead of unified targets. For the past fifty years since Rai and Binet staging of CLL was first described (Binet et al., 1981; Rai et al., 1975), holistic patient characteristics such as age and disease burden have continued to contribute to disease staging (Bahlo et al., 2016), suggesting these factors may confound efforts to deconvolute genetics-specific effects in patients. Furthermore, genetic mouse knock-out models have knocked out portions of del(11) and del(13) in every cell of the mouse (Klein et al., 2010; Yin et al., 2019), making the conclusions and targets difficult to confidently apply to human patients who experience these deletions only in B cells. In any case, tri12 is embryonically lethal (Hassold et al., 1980), so a traditional tri12 mouse model is unavailable. With these limitations to patient samples and mouse models, my human PSCs tri12 offered a new way to investigate CLL and an outsider's perspective of the field.

I therefore set out to investigate tri12 in human PSCs, with an isogenic counterpart, to identify actionable targets to inhibit malignant tri12 CLL. Through my *in vivo* co-transplant experimentation efforts, I developed a novel assay to test the selective effects of compounds in xenotransplant mice; injected tri12 and isogenic PSCs to form tumors in the same mouse in separate locations (***Chapter 2***). To my knowledge, this concept has not been previously reported. Instead of *in vivo* co-transplantation selectivity models, our lab (Lee et al., 2017b), and others (Kotini et al., 2017), have created PSC lines from acute myeloid leukemia (AML) patients. Together with my colleagues, we reported that BET inhibitors selectively killed AML iPSCs *in vitro*, but did not affect isogenic iPSCs (Lee et al., 2017b). This drug-gene link had been previously shown *in vivo* (Fong et al., 2015). Co-transplantation of mutant and isogenic iPSC lines into the same mouse may open a new avenue of preclinical testing for genetics-associated diseases, especially for diseases and tissue types that are more difficult to model *in vivo*. This could be applied in many fields beyond malignant hematopoiesis.

5.2.2 Future clinical perspectives: novel targets for precision oncology

Selective molecular targeting has become a successful standard-of-care in leukemia; such as chronic myelogenous leukemia (BCR-ABL1) and acute promyelocytic leukemia (PML-RARA). In the most common adult hematologic malignancy, CLL, patients have not benefited from similar molecular-targeting therapies arising from genetic mutations. Historically, first-in-class non-chemotherapy drugs for CLL are developed using lymphoma cell lines in mice and then attempted as single doses in relapse/refractory CLL patients (Souers et al., 2013), neither of which is useful for tri12-specific genes targeted repeatedly. Clinicians are quite excited about the adoption of non-chemotherapy for CLL treatment which transcends genetics (van der Straten et al., 2020), and rightly so, however these 'B cell-identity targeted therapies' produce many severe off-target effects which still need to be controlled and improved upon (Brown, 2018), suggesting combination therapies together with further targeted therapies might provide a means for dose reductions and reduced toxicity while achieving improved efficacy. Furthermore, the two lead drugs I developed for tri12

selectivity appear to be tolerated very well *in vivo* for extended durations and are orally bioavailable, both desirable dosing properties.

As a next step, it will be important to regulate the gene dosage directly of lead candidate genes and/or inhibit the entirety of the third chromosome 12 to demonstrate a causal role. Experimental RNA silencing (shRNA or siRNA) would provide more direct evidence to support the claims of the specificity of the targets I identified. Furthermore, experimental studies have shown that gene dosage resulting from a trisomy can be improved using transgenic expression of *XIST* on the offending chromosome; the gene used in development to silence the extra X chromosome of females by repressing all gene expression on the entire chromosome (Czerminski and Lawrence, 2020). Ectopic expression of *XIST* on one of the three chromosome 12's may alter the pathology of tri12 CLL and improve our understanding of it. These possibilities should be further explored.

5.3 Unique properties of hematopoietic development from PSCs

There exists a dizzying number and complexity of differentiation protocols for producing various hematopoietic cells types from human PSCs, including fully differentiated cells as well as hematopoietic progenitors and even earlier endothelial precursor cells. Due to space constraints, I apologize to all the researchers who labored to modify and improve upon these protocols, whom I cannot include in this brief summary as I contextualize my results.

5.3.1 Future translational perspectives: lymphoid development

Myelo-erythroid HPCs are readily and reproducibly produced from human PSCs (Chadwick et al., 2003; Vodyanik et al., 2005). Technical improvements adapting these protocols have increased the efficiency of producing hPSC-derived hPSCs to produce macrophages and orthochromatic normoblasts which are morphologically similar to somatic HPCs (Reid et al., 2018). Embryonic primitive hematopoiesis produces erythrocytes, macrophages and megakaryocytes, but does not produce HSCs and B and T lymphoid cells (Xu et al., 2001). Therefore, there exists the hypothesis that hPSC-HSCs can't be differentiated *in vitro* because the resulting cells are 'embryonic' in nature. Evidence to support this claim comes from the difficulty in deriving lymphoid cells as well as embryonic globin genes expressed in orthochromatic normoblasts; primitive erythrocytes which do not enucleate (McIntyre et al., 2013). However, multiple studies have reported production of natural killer cells (Vodyanik et al., 2005), T lymphoid differentiation (Ditadi et al., 2015), and most relevant to my work: B lymphoid differentiation (Carpenter et al., 2011; French et al., 2015; Galat et al., 2017; Vodyanik et al., 2005). However, when I was learning to produce myelo-erythroid cells, the simple cocktail and protocol established by our lab (Chadwick et al., 2003), was always successful and highly reproducible between researchers and across PSC lines and batches of reagents, and produced functional myelo-erythroid progenitors within two weeks. In contrast, it took me a year and multiple stages of optimization of cell stoichiometry, media composition, and durations of differentiation, to produce an abundance of B cells, but only from tri12 PSCs, as the same protocol completed in parallel with isogenic PSC remained inefficient but still possible to achieve a few phenotypic B cells. Therefore, these data suggest that the original hypothesis made sense at the time when the default myelo-erythroid cells

were easier to produce, but multiple reports have now cast doubt as to whether the inability to produce HSCs is because of an embryonic program.

Since lymphoid development of hPSCs is possible, this enables the modelling of B lymphoid disease including CLL. Primary CLL cells do not proliferate *in vitro* or *in vivo* with current protocols, making them a finite resource, while immortalized cell lines from CLL patients lose key features of the pathobiology (Stacchini et al., 1999; Wendel-Hansen et al., 1994), and are complex karyotype (Kellner et al., 2016; Stacchini et al., 1999) which is not commonly observed in patients (Döhner et al., 2000). Beyond CLL, antigen presenting cells are important in the modern reality that CAR-T cell technology is biologically effective but prohibitively expensive to use broadly. Researchers are already calling for 'off the shelf' T cells from pluripotent sources (Iriguchi and Kaneko, 2019). In addition to safety and efficacy, the development of B and T lymphoid requires further experimentation to move beyond mouse stromal cell lines representing the only the BM. B and T lymphoid cells both go through maturation processes outside of the BM (i.e. lymph nodes, spleen, and thymus), and may even require each other to fully mature as well (Clark and Ledbetter, 1994). Therefore, it seems apparent that differentiation procedures should screen cues relevant to these organs and assess more than just phenotype at the end of each protocol.

5.3.2 Future translational perspectives: in vivo hematopoietic development

The field has remained diversified in the approach to produce hPSC-derived HSCs, but remains united in its goal to achieve SRC capacity; multi-lineage human reconstitution in mouse xenografts (Bhatia et al., 1997). From the early days with multiple cytokines and growth factors continuously available in media (Chadwick et al., 2003), to step-wise addition of these same factors (Lee et al., 2017a), co-culture with mouse stromal cells (Vodyanik et al., 2005), to the inclusion of lentivirus delivery of transcription factors to promote hematopoietic development (Doulatov et al., 2013; Sugimura et al., 2017), all have demonstrated incremental improvements in HPC quantity but not to the point of SRC capacity comparable to somatic sources such as adult BM or umbilical CB. Despite the modest successes and major advances, this field of research holds potential in basic and translational research. Early on, transcription factors which 'create' HSCs from mouse PSCs (Kyba et al., 2002), could not be similarly demonstrated using human PSCs (Wang, 2005). This emphasized the necessity to study human HSCs and not mouse surrogates.

Informed upon this field of research, my work to ectopically express CXCR4 in hPSC-HPCs and test SRC capacity with functional CFU output after xenotransplantation was also an incremental achievement (Reid et al., 2018). Based on my colleagues' work with healthy HSCs and primary leukemia samples, it soon became evident to me that these cells are quite dependent on their BM microenvironment for cues that included survival, location, and maturity (Boyd et al., 2014; Boyd et al., 2017; Guezguez et al., 2013). At the same time, multiple reports of transcription factors used to alter hPSC-HPCs were being reported (Doulatov et al., 2013; Ramos-Mejia et al., 2014; Ran et al., 2013). I therefore *hypothesized that proteins which enable regulation of HSC survival in the BM promote the survival and function of human PSC-derived HPCs.* The paucity of evidence demonstrating hPSC-HPCs interact with their microenvironment, *in vitro* and *in vivo*, was an appealing opportunity I wanted to pursue. In turn, studying the microenvironment cues of hPSC-HPCs may inform upon what is necessary and sufficient for somatic hematopoietic cells. It is evident that CXCR4

is expressed on most hematopoietic cells, while somatic CX3CR1 expression is predominately restricted to macrophages in the hematopoietic system (Imai, 1997). These receptors mediate cellular adhesion as well as downstream signaling in hematopoietic cells. My work with CXCR4 shows that it has not been functioning for hPSC-HPCs and can be used to increase progenitor output (Reid et al., 2018). I think it is quite intriguing that most cell surface markers of primitive HSCs are cellular adhesion molecules which interact with the microenvironment; CD34, CD49f (*ITGA6*), CD90 (Thy-1), GPI-80 (*VNN2*). Knock-out studies of these proteins have not established their essential roles in HSCs; i.e. CD34 knock out mice (Cheng et al., 1996). However, this includes the inherent limitations of knockout studies in mice, which are significant for the study of HSCs since HSC markers in humans and mice are quite different. Receptor overlap and redundancies to protect and exquisitely regulate HSCs *in situ* are also intuitively reported (Mancini et al., 2005). Perhaps over-expression studies of these proteins in a null background, such as hPSC-HPCs, could provide an alternative means to determine the biological properties of these proteins and their fundamental roles in BM microenvironmental interactions *in vivo*.

5.4 Unique properties of healthy hematopoietic regeneration

The regeneration of hematopoietic and non-hematopoietic cells of the BM following myelosuppression has received extensive interest by our group (Boyd et al., 2018) and others (Silberstein et al., 2016; Tikhonova et al., 2019; Zheng et al., 2017). Furthermore, agents such as angiogenin (Goncalves et al., 2016) and cell cycle inhibitors (He et al., 2017), have been developed to reduce 'HSC exhaustion' following CIM, but this approach may not meet the short term needs of patients with acute myelosuppression. Furthermore, studies commonly study only one form of cytotoxic agent, and there are at least three commonly used myelosuppressive agents: radiation (Goncalves et al., 2016), 5-FU (He et al., 2017), and AraC in the context of leukemia (Boyd et al., 2018), making it difficult to draw comparisons across studies. Despite single cell resolution of non-hematopoietic cells of the mouse BM pre- and post-myelosuppression, this information is incomplete for human hematopoietic cells. Furthermore, there is a paucity of long-term follow-up and functionally defined time points of regeneration, making it difficult to compare results across studies with different forms of myelosuppression as well as different time points analyzed. This lead me to evaluate multiple forms of myelosuppression and multiple time points of recovery to ask whether there *are key cells and pathways conserved during hematopoietic recovery from myelosuppressive injuries.*

5.4.1 Future clinical perspectives: endogenous regeneration

This work began as an open ended follow-up to the definition of an HRC signature in the context of a 'negative control' for the LRC signature (Boyd et al., 2018). The cell types and signaling pathways present in the HRC state were unknown, as well as the underlying biology of recovery which suggested healthy regeneration was extremely limited compared to primary AML xenografted mice. These observations lead me to ask when does healthy regeneration occur? What will be the measure of this? I developed several models of myelosuppression in addition to AraC and contrasted them phenotypically and functionally at multiple regularly interspaced time points after myelosuppressive treatment. I found similarities in the underlying biology, which I then assessed more deeply using single cell

RNA-seq. Importantly, these new 'HRPC' time points could not have been developed in an isolated cell culture dish, since the BM releases complex cascades of signals to stimulate repair as well (Silberstein et al., 2016). Indeed, I did not observe up-regulation of HRC markers when healthy progenitors were exposed to serum isolated from CIM-treated mice, while I reproducibly observed this when healthy progenitors were purified from CIM-treated mice. Usually, I prefer the simplest model that can describe the phenomenon, but in this circumstance, I could not have developed the conclusions I did by using *in vitro* models.

To further expand on the biology of HRPC and HRC *in situ* following CIM, it will be important to evaluate these markers in the context of multiple rounds of CIM. Many patients receive multiple cycles of chemotherapy. Do HRC markers resolve prior to the next round of CIM? Are they further upregulated or become dulled responses? Does exhaustion observed in repeated limb amputation in axolotls (Bryant et al., 2017) correlate to the biology observed with repeated CIM? Using the long-term follow-up HRC model in human-mouse xenografts is well suited to address these questions (***Chapter 4***), as repeated BM aspiration enables the assessment of chimerism at multiple time points, as opposed to performing secondary, tertiary, quaternary, *etc*., xenotransplants to assess residual HSPCs. Serial transplantation also removes compounded effects which would represent a significant loss since the same patient receives multiple rounds of CIM

The nature of the composition of each whole BM human-mouse xenograft following CIM demands investigations as well. I discovered primitive cells, neutrophils, dendritic cells, and macrophages became a larger proportion of the graft at HRPC time points (***Chapter 4***). However, it remains unclear whether these populations are expanded through proliferation or because of a diminishment of another compartment, such as B lymphoid cells. Since I have counted the BM and assessed the total chimerism and many hematopoietic lineages using flow cytometry, I observe all compartments becoming repressed and that primitive and myeloid cells expand in total counts within the BM. However, these data are correlative not causative. I'm not confident genetic barcoding systems would be useful in the context of extensively engrafted hosts prior to CIM, since these barcodes would be fully throughout the entire hematopoietic system in multiple types of progenitors by 7 weeks. However, classic lineage tracing techniques such as Cre-loxP recombination (Hsu, 2015), triggered in progenitor populations by an exogenous drug at the time of CIM, would permit the study of expanded cell populations in a more causal manner. Indeed, these possibilities will inform upon the nature of an interventional approach to stimulate recovery endogenously.

5.4.2 Future clinical perspectives: healthy hematopoiesis in leukemia patients

Since the HRC signature was developed specifically to highlight divergent pathways not enriched in LRC (Boyd et al., 2018), this formed the basis to investigate what these pathways and cell identities are so that they could potentially be useful for AML patients. Although my thesis does not revolve around AML, my co-authored works with Dr. Allison Boyd have taught me that healthy hematopoiesis is impaired in the presence of AML (Boyd et al., 2017), and that supporting the healthy component is critical since the lack of these cells increases patient morbidity. The HRC model is fundamentally based on the premise that the healthy hematopoietic system is not intentionally being damaged but is a necessary evil in order to provide lifesaving chemotherapy for alternative reasons; i.e. breast and colon cancer (5-FU) or leukemia (AraC). It is especially important to note that myeloid expansion (with the

caveats discussed in the previous section) and in particular macrophage expansion was observed following healthy regeneration, whereas AML is associated with differentiation blockade of multiple types of myelo-erythroid cells (Boyd et al., 2017). Tumor-associated macrophages are associated with tumor progression, and promote angiogenesis in solid tumors, and their origin is unclear (Yang et al., 2018). Further molecular characterization to determine whether HRC macrophages perform clean-up of the microenvironment (called 'M1') and/or exacerbate tumor-like growth (called 'M2'), is warranted (Yang et al., 2018). I hypothesize it is the former, given *VASH1* was highly expressed in HRPC macrophages, and vasohibin1 (*VASH1* gene product) is an inhibitor of angiogenesis and migration (Hinamoto et al., 2014). Indeed, *VASH1* knock out mice lack distinct macrophage subtypes and accumulate renal fibrosis in diabetes-induced injury. CIM in the case of co-transplantation of both healthy and AML primary samples in xenotransplant mice (using divergent HLA-A2 expression; Boyd et al. 2014) would cleanly evaluate macrophage composition (and *VASH1* expression). I would expect significant differences given we previously reported inhibitory effects of leukemia on healthy maturation using a co-transplant xenograft model (Boyd et al., 2017). Transfusions of healthy macrophages into singly transplanted leukemic xenografts as well as co-transplanted xenografts is warranted to determine whether impaired resolution of myelosuppression in leukemia patients with myeloid differentiation blockade (Inagawa et al., 2019), is due to deficiencies in healthy HRC capacity which can be improved with cellular products, related growth factors, or chemical agents.

5.5 Concluding remarks

This thesis represents a collection of work which sought to deconvolute healthy and transformed human hematopoiesis using multiple novel cell-based assays and investigative tools. Through the varied use of primary cell cultures, lentiviral gene therapeutics, xenotransplant models, single cell RNA-sequencing, machine learning, and further bioinformatics methods, I have been able to explore multiple fascinating aspects of hematopoietic regulation, specifically leukemic mutation-based transformation and healthy progenitor cell behaviour. As alluded to throughout this thesis, these models (e.g. isogenic-controlled tri12 PSC) shed light on an untapped area of research in leukemic research. It was the goal of this work to help focus the importance of function-based mechanisms and experimentation in order to develop meaningful insights. I first developed this approach with lentiviral gene therapeutics to enhance hPSC-HPCs, which is apparent in my later work using human-mouse xenografts and treating the mouse with chemotherapy to dissect the systemic effects of CIM on human hematopoiesis *in vivo*. With ongoing efforts to adapt isogenic-controlled PSC elucidation of leukemic mutations and developing applications for HRPC biology, future efforts in the field of regenerative hematopoiesis will no doubt yield exciting results.

Appendix

Appendix I: Bibliography

Ahn, I.E., Underbayev, C., Albitar, A., Herman, S.E.M., Tian, X., Maric, I., Arthur, D.C., Wake, L., Pittaluga, S., Yuan, C.M.*, et al.* (2017). Clonal evolution leading to ibrutinib resistance in chronic lymphocytic leukemia. Blood *129*, 1469-1479.

AlJanahi, A.A., Danielsen, M., and Dunbar, C.E. (2018). An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. Mol Ther Methods Clin Dev *10*, 189-196.

Amabile, G., Welner, R.S., Nombela-Arrieta, C., D'Alise, A.M., Di Ruscio, A., Ebralidze, A.K., Kraytsberg, Y., Ye, M., Kocher, O., Neuberg, D.S.*, et al.* (2013). In vivo generation of transplantable human hematopoietic cells from induced pluripotent stem cells. Blood *121*, 1255-1264.

Anonymous (2010). Human embryonic stem cell research in the US: time for change? Nat Cell Biol *12*, 627.

Aslostovar, L., Boyd, A., Almakadi, M., Collins, T., Leong, D., Tirona, R., Kim, R., Julian, J., Xenocostas, A., Leber, B.*, et al.* (2018). A phase 1 trial evaluating thioridazine in combination with cytarabine in patients with acute myeloid leukemia. Blood Advances *2*, 1935-1945.

Bahlo, J., Kutsch, N., Bauer, K., Bergmann, M.A., Byrd, J.C., Chaffee, K.G., Döhner, H., Eichhorst, B., Else, M., Geisler, C.*, et al.* (2016). An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. The Lancet Oncology *17*, 779-790.

Bain, B.J. (2001). Bone marrow trephine biopsy. Journal of Clinical Pathology *54*, 737.

Banno, K., Omori, S., Hirata, K., Nawa, N., Nakagawa, N., Nishimura, K., Ohtaka, M., Nakanishi, M., Sakuma, T., Yamamoto, T.*, et al.* (2016). Systematic Cellular Disease Models Reveal Synergistic Interaction of Trisomy 21 and GATA1 Mutations in Hematopoietic Abnormalities. Cell reports.

Baron, C.S., Barve, A., Muraro, M.J., van der Linden, R., Dharmadhikari, G., Lyubimova, A., de Koning, E.J.P., and van Oudenaarden, A. (2019). Cell Type Purification by Single-Cell Transcriptome-Trained Sorting. Cell *179*, 527-542 e519.

Batsivari, A., Haltalli, M.L.R., Passaro, D., Pospori, C., Lo Celso, C., and Bonnet, D. (2020). Dynamic responses of the haematopoietic stem cell niche to diverse stresses. Nature Cell Biology *22*, 7-17.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nature biotechnology *37*, 38-44.

Becker, A.J., McCulloch, E.A., and Till, J.E. (1963). Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature *197*, 452-454.

Beekman, R., Chapaprieta, V., Russiñol, N., Vilarrasa-Blasi, R., Verdaguer-Dot, N., and Martens, J. (2018). The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nature Medicine.

Bhatia, M., Wang, J.C.Y., Kapp, U., Bonnet, D., and Dick, J.E. (1997). Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. Proceedings of the National Academy of Sciences of the United States of America *94*, 5320-5325.

Bichi, R., Shinton, S.A., Martin, E.S., Koval, A., Calin, G.A., Cesari, R., Russo, G., Hardy, R.R., and Croce, C.M. (2002). Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression. Proceedings of the National Academy of Sciences of the United States of America *99*, 6955-6960.

Binet, J., Auquier, A., Dighiero, G., Chastang, C., Piguet, H., Goasguen, J., Vaugier, G., Potron, G., Colona, P., Oberling, F.*, et al.* (1981). A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer *48*, 198-206.

Birbrair, A., and Frenette, P.S. (2016). Niche heterogeneity in the bone marrow. Annals of the New York Academy of Sciences *1370*, 82-96.

Bosma, G.C., Custer, R.P., and Bosma, M.J. (1983). A severe combined immunodeficiency mutation in the mouse. Nature *301*, 527-530.

Boyd, A., Aslostovar, L., Reid, J., Ye, W., Tanasijevic, B., Porras, D., Shapovalova, Z., Almakadi, M., Foley, R., Leber, B.*, et al.* (2018). Identification of chemotherapy-induced Leukemic Regenerating Cells reveals a transient vulnerability of human AML recurrence. Cancer Cell *34*, 483-498.

Boyd, A.L., Campbell, C.J., Hopkins, C.I., Fiebig-Comyn, A., Russell, J., Ulemek, J., Foley, R., Leber, B., Xenocostas, A., Collins, T.J.*, et al.* (2014). Niche displacement of human leukemic stem cells uniquely allows their competitive replacement with healthy HSPCs. The Journal of experimental medicine *211*, 1925-1935.

Boyd, A.L., Reid, J.C., Salci, K.R., Aslostovar, L., Benoit, Y.D., Shapovalova, Z., Nakanishi, M., Porras, D.P., Almakadi, M., Campbell, C.J.V.*, et al.* (2017). Acute myeloid leukaemia disrupts endogenous myelo-erythropoiesis by compromising the adipocyte bone marrow niche. Nat Cell Biol.

Brenet, F., Kermani, P., Spektor, R., Rafii, S., and Scandura, J.M. (2013). TGFbeta restores hematopoietic homeostasis after myelosuppressive chemotherapy. The Journal of experimental medicine *210*, 623-639.

Brieghel, C., da Cunha-Bang, C., Yde, C.W., Schmidt, A.Y., Kinalis, S., Nadeu, F., Andersen, M.A., Jacobsen, L.O., Andersen, M.K., Pedersen, L.B.*, et al.* (2020). The Number of Signaling Pathways Altered by Driver Mutations in Chronic Lymphocytic Leukemia Impacts Disease Outcome. Clinical Cancer Research, clincanres.4158.2018.

Brown, J. (2018). How I treat CLL patients with ibrutinib. Blood *131*, 379-386.

Broxmeyer, H.E., Douglas, G.W., Hangoc, G., Cooper, S., Bard, J., English, D., Arny, M., Thomas, L., and Boyse, E.A. (1989). Human umbilical cord blood as a potential source of transplantable hematopoietic stem/progenitor cells. Proceedings of the National Academy of Sciences of the United States of America *86*, 3828-3832.

Broxmeyer, H.E., Mejia, J.A., Hangoc, G., Barese, C., Dinauer, M., and Cooper, S. (2007). SDF-1/CXCL12 enhances in vitro replating capacity of murine and human multipotential and macrophage progenitor cells. Stem cells and development *16*, 589-596.

Broxmeyer, H.E., Orschell, C.M., Clapp, D.W., Hangoc, G., Cooper, S., Plett, P.A., Liles, W.C., Li, X., Graham-Evans, B., Campbell, T.B.*, et al.* (2005). Rapid mobilization of murine and human hematopoietic stem and progenitor cells with AMD3100, a CXCR4 antagonist. The Journal of experimental medicine *201*, 1307-1318.

Bryant, D.M., Sousounis, K., Payzin-Dogru, D., Bryant, S., Sandoval, A.G.W., Martinez Fernandez, J., Mariano, R., Oshiro, R., Wong, A.Y., Leigh, N.D.*, et al.* (2017). Identification of regenerative roadblocks via repeat deployment of limb regeneration in axolotls. Regenerative medicine *2*.

Bujko, K., Cymer, M., Adamiak, M., and Ratajczak, M. (2019). An Overview of Novel Unconventional Mechanisms of Hematopoietic Development and Regulators of Hematopoiesis – a Roadmap for Future Investigations. Stem Cell Reviews and Reports.

Carpenter, L., Malladi, R., Yang, C.T., French, A., Pilkington, K.J., Forsey, R.W., Sloane-Stanley, J., Silk, K.M., Davies, T.J., Fairchild, P.J.*, et al.* (2011). Human induced pluripotent stem cells are capable of B-cell lymphopoiesis. Blood *117*, 4008-4011.

Chadwick, K., Wang, L., Li, L., Menendez, P., Murdoch, B., Rouleau, A., and Bhatia, M. (2003). Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. Blood *102*, 906-915.

Cheng, J., Baumhueter, S., Cacalano, G., Carver-Moore, K., Thibodeaux, H., Thomas, R., Broxmeyer, H.E., Cooper, S., Hague, N., Moore, M.*, et al.* (1996). Hematopoietic defects in mice lacking the sialomucin CD34. Blood *87*, 479-490.

Choi, K.-D., Vodyanik, M., and Slukvin, I.I. (2011). Hematopoietic differentiation and production of mature myeloid cells from human pluripotent stem cells. Nat Protoc *6*, 296-313.

Choi, M.R., Solid, C.A., Chia, V.M., Blaes, A.H., Page, J.H., Barron, R., and Arneson, T.J. (2014). Granulocyte colony-stimulating factor (G-CSF) patterns of use in cancer patients receiving myelosuppressive chemotherapy. Support Care Cancer *22*, 1619-1628.

Clark, E.A., and Ledbetter, J.A. (1994). How B and T cells talk to each other. Nature *367*, 425-428.

Clough, E., and Barrett, T. (2016). The Gene Expression Omnibus Database. Methods in molecular biology (Clifton, NJ) *1418*, 93-110.

Cutler, C., Multani, P., Robbins, D., Kim, H.T., Le, T., Hoggatt, J., Pelus, L.M., Desponts, C., Chen, Y.B., Rezner, B.*, et al.* (2013). Prostaglandin-modulated umbilical cord blood hematopoietic stem cell transplantation. Blood *122*, 3074-3081.

Czerminski, J.T., and Lawrence, J.B. (2020). Silencing Trisomy 21 with XIST in Neural Stem Cells Promotes Neuronal Differentiation. Dev Cell.

De Sousa, P.A., Tye, B.J., Bruce, K., Dand, P., Russell, G., Collins, D.M., Greenshields, A., McDonald, K., Bradburn, H., Laurie, A.*, et al.* (2016). Derivation of the clinical grade human embryonic stem cell line RCe017-A (RC-13). Stem cell research *16*, 756-760.

Ditadi, A., Sturgeon, C.M., Tober, J., Awong, G., Kennedy, M., Yzaguirre, A.D., Azzola, L., Ng, E.S., Stanley, E.G., French, D.L.*, et al.* (2015). Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. Nat Cell Biol *17*, 580-591.

Döhner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Kröber, A., Bullinger, L., Döhner, K., and Bentz, M., Lichter, P. (2000). Genomic aberrations and survival in chronic lymphocytic leukemia. The New England Journal of Medicine *343*, 1910-1916.

Doulatov, S., Vo, L., Chou, S., Kim, P., Arora, N., Li, H., Hadland, B., Bernstein, I., Collins, J., Zon, L.*, et al.* (2013). Induction of Multipotential Hematopoietic Progenitors from Human Pluripotent Stem Cells via Respecification of Lineage-Restricted Precursors. Cell stem cell *13*, 459-470.

Doulatov, S.N., F.; Laurenti, E.; Dick, J. E. (2012). Hematopoiesis: a human perspective. Cell stem cell *10*, 120-136.

Draper, J.S., Smith, K., Gokhale, P., Moore, H.D., Maltby, E., Johnson, J., Meisner, L., Zwaka, T.P., Thomson, J.A., and Andrews, P.W. (2003). Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. Nature biotechnology *22*, 53-54.

Druker, B., Tamura, S., Buchdunger, E., Ohno, S., Segal, G., Fanning, S., Zimmermann, J., and Lydon, N. (1996). Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr–Abl positive cells. Nature Medicine, 561-566.

Dührsen, U., Villeval, J.L., Boyd, J., Kannourakis, G., Morstyn, G., and Metcalf, D. (1988). Effects of recombinant human granulocyte colony-stimulating factor on hematopoietic progenitor cells in cancer patients. Blood *72*, 2074-2081.

Durig, J., Ebeling, P., Grabellus, F., Sorg, U.R., Mollmann, M., Schutt, P., Gothert, J., Sellmann, L., Seeber, S., Flasshove, M.*, et al.* (2007). A novel nonobese diabetic/severe combined immunodeficient xenograft model for chronic lymphocytic leukemia reflects important clinical characteristics of the disease. Cancer Res *67*, 8653-8661.

Eaves, C., Miller, C., Conneally, E., Audet, J., Oostendorp, R., Cashman, J., Zandstra, P., Rose-John, S., Piret, J., and Eaves, A. (1999). Introduction to stem cell biology in vitro. Threshold to the future. In Ann N Y Acad Sci, pp. 1-8.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Fares, I., Chagraoui, J., Lehnertz, B., MacRae, T., Mayotte, N., Tomellini, E., Aubert, L., Roux, P.P., and Sauvageau, G. (2017). EPCR expression marks UM171-expanded CD34+ cord blood stem cells. Blood *129*, 3344-3351.

Fischer, K., Al-Sawaf, O., Bahlo, J., Fink, A.M., Tandon, M., Dixon, M., Robrecht, S., Warburton, S., Humphrey, K., Samoylova, O.*, et al.* (2019). Venetoclax and Obinutuzumab in Patients with CLL and Coexisting Conditions. N Engl J Med *380*, 2225-2236.

Fong, C.Y., Gilan, O., Lam, E.Y., Rubin, A.F., Ftouni, S., Tyler, D., Stanley, K., Sinha, D., Yeh, P., Morison, J.*, et al.* (2015). BET inhibitor resistance emerges from leukaemia stem cells. Nature *525*, 538-542.

Ford, C.E., Hamerton, J.L., Barnes, D.W., and Loutit, J.F. (1956). Cytological identification of radiation-chimaeras. Nature *177*, 452-454.

Fraser, G., Cramer, P., Demirkan, F., Silva, R., Grosicki, S., Pristupa, A., Janssens, A., Mayer, J., Bartlett, N., Dilhuydy, M.*, et al.* (2018). Updated results from the phase 3 HELIOS study of ibrutinib, bendamustine, and rituximab in relapsed chronic lymphocytic leukemia/small lymphocytic lymphoma. Leukemia.

French, A., Yang, C.T., Taylor, S., Watt, S.M., and Carpenter, L. (2015). Human induced pluripotent stem cell-derived B lymphocytes express sIgM and can be generated via a hemogenic endothelium intermediate. Stem cells and development *24*, 1082-1095.

Galat, Y., Dambaeva, S., Irina Elcheva, I., Khanolkar, A., Beaman, K., Iannaccone, P., and Galat, V. (2017). Cytokine-free directed differentiation of human pluripotent stem cells efficiently produces hemogenic endothelium with lymphoid potential. Stem Cell Res Ther *8*.

Gertow, K., Cedervall, J., Unger, C., Szöke, K., Blennow, E., Imreh, M.P., and Ährlund-Richter, L. (2007). Trisomy 12 in HESC leads to no selective in vivo growth advantage in teratomas, but induces an increased abundance of renal development. J Cell Biochem *100*, 1518-1525.

Goncalves, K.A., Silberstein, L., Li, S., Severe, N., Hu, M.G., Yang, H., Scadden, D.T., and Hu, G.F. (2016). Angiogenin Promotes Hematopoietic Regeneration by Dichotomously Regulating Quiescence of Stem and Progenitor Cells. Cell *166*, 894-906.

Green, D.E., and Rubin, C.T. (2014). Consequences of irradiation on bone and marrow phenotypes, and its relation to disruption of hematopoietic precursors. Bone *63*, 87-94.

Green, E.D., Watson, J.D., and Collins, F.S. (2015). Human Genome Project: Twenty-five years of big biology. Nature *526*, 29-31.

Guezguez, B., Campbell, C.J., Boyd, A.L., Karanu, F., Casado, F.L., Di Cresce, C., Collins, T.J., Shapovalova, Z., Xenocostas, A., and Bhatia, M. (2013). Regional localization within the bone marrow influences the functional capacity of human HSCs. Cell stem cell *13*, 175-189.

Guo, B., Huang, X., Cooper, S., and Broxmeyer, H.E. (2017). Glucocorticoid hormone-induced chromatin remodeling enhances human hematopoietic stem cell homing and engraftment. Nat Med.

Hassold, T., Chen, N., Funkhouser, J., Jooss, T., Manuel, B., Matsuura, J., Matsuyama, A., Wilson, C., Yamane, J., and Jacobs, P. (1980). A cytogenetic study of 1000 spontaneous abortions. Ann Hum Genet *44*, 151-178.

He, S., Roberts, P.J., Sorrentino, J.A., Bisi, J.E., Storrie-White, H., Tiessen, R.G., Makhuli, K.M., Wargin, W.A., Tadema, H., van Hoogdalem, E.J.*, et al.* (2017). Transient CDK4/6 inhibition protects hematopoietic stem cells from chemotherapy-induced exhaustion. Sci Transl Med *9*.

Healy, L., May, G., Gale, K., Grosveld, F., Greaves, M., and Enver, T. (1995). The stem cell antigen CD34 functions as a regulator of hemopoietic cell adhesion. Proceedings of the National Academy of Sciences of the United States of America *92*, 12240-12244.

Hentze, H., Soong, P.L., Wang, S.T., Phillips, B.W., Putti, T.C., and Dunn, N.R. (2009). Teratoma formation by human embryonic stem cells: Evaluation of essential parameters for future safety studies. In Stem cell research (Elsevier B.V.), pp. 198-210.

Herault, A., Binnewies, M., Leong, S., Calero-Nieto, F.J., Zhang, S.Y., Kang, Y.A., Wang, X., Pietras, E.M., Chu, S.H., Barry-Holson, K.*, et al.* (2017). Myeloid progenitor cluster formation drives emergency and leukaemic myelopoiesis. Nature *544*, 53-58.

Hinamoto, N., Maeshima, Y., Yamasaki, H., Nasu, T., Saito, D., Watatani, H., Ujike, H., Tanabe, K., Masuda, K., Arata, Y.*, et al.* (2014). Exacerbation of diabetic renal alterations in mice lacking vasohibin-1. PloS one *9*, e107934.

Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V.*, et al.* (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell *173*, 291-304 e296.

Honigberg, L., Smith, A., Sirisawad, M., Verner, E., Loury, D., Chang, B., Li, S., Pan, Z., Thamm, D., Miller, R.*, et al.* (2010). The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. Proceedings of the National Academy of Sciences of the United States of America *107*, 13075-13080.

Hsu, Y.-C. (2015). Theory and Practice of Lineage Tracing. Stem cells (Dayton, Ohio) *33*, 3197-3204.

Hu, B., Patel, K.P., Chen, H.C., Wang, X., Luthra, R., Routbort, M.J., Kanagal-Shamanna, R., Medeiros, L.J., Yin, C.C., Zuo, Z.*, et al.* (2019). Association of gene mutations with time-to-first treatment in 384 treatment-naive chronic lymphocytic leukaemia patients. Br J Haematol.

Ichikawa, Y., Pluznik, D.H., and Sachs, L. (1966). In vitro control of the development of macrophage and granulocyte colonies. Proceedings of the National Academy of Sciences of the United States of America *56*, 488-495.

Imai, T.H., K.; Haskell, C.; Baba, M.; Nagira, M.; Nishimura, M.; Kakizaki, M.; Takagi, S.; Nomiyama, H.; Schall, T.; Yoshie, O. (1997). Identification and Molecular Characterization of Fractalkine Receptor CX3CR1, which Mediates Both Leukocyte Migration and Adhesion. Cell *91*, 521-530.

Inagawa, Y., Komeno, Y., Saito, S., Maenohara, Y., Yamagishi, T., Kawashima, H., Saito, T., Abe, K., Iihara, K., Hatada, Y.*, et al.* (2019). Prolonged Myelosuppression due to Progressive Bone Marrow Fibrosis in a Patient with Acute Promyelocytic Leukemia. Case Rep Hematol *2019*, 1616237.

Iriguchi, S., and Kaneko, S. (2019). Toward the development of true "off-the-shelf" synthetic T-cell immunotherapy. Cancer Sci *110*, 16-22.

Ito, R., Takahashi, T., Katano, I., and Ito, M. (2012). Current advances in humanized mouse models. Cellular and Molecular Immunology *9*, 208-214.

Jacobson, L.O., Simmons, E.L., Marks, E.K., and Eldredge, J.H. (1951). Recovery from radiation injury. Science (New York, NY) *113*, 510-511.

Kellner, J., Wierda, W., Shpall, E., Keating, M., and McNiece, I. (2016). Isolation of a novel chronic lymphocytic leukemic (CLL) cell line and development of an in vivo mouse model of CLL. Leuk Res *40*, 54-59.

Kennedy, M., Awong, G., Sturgeon, C.M., Ditadi, A., LaMotte-Mohs, R., Zuniga-Pflucker, J.C., and Keller, G. (2012). T lymphocyte potential marks the emergence of definitive hematopoietic progenitors in human pluripotent stem cell differentiation cultures. Cell reports *2*, 1722-1735.

Kimura, H., Burstein, S.A., Thorning, D., Powell, J.S., Harker, L.A., Fialkow, P.J., and Adamson, J.W. (1984). Human megakaryocytic progenitors (CFU-M) assayed in methylcellulose: Physical characteristics and requirements for growth. Journal of Cellular Physiology *118*, 87-96.

Klein, U., Lia, M., Crespo, M., Siegel, R., Shen, Q., Mo, T., Ambesi-Impiombato, A., Califano, A., Migliazza, A., Bhagat, G.*, et al.* (2010). The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. Cancer Cell *17*, 28-40.

Kotini, A.G., Chang, C.-J., Chow, A., Yuan, H., Ho, T.-C., Wang, T., Vora, S., Solovyov, A., Husser, C., Olszewska, M.*, et al.* (2017). Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia. Cell stem cell *20*, 1-14.

Kyba, M., Perlingeiro, R., and Daley, G.Q. (2002). HoxB4 Confers Definitive Lymphoid-Myeloid Engraftment Potential on Embryonic Stem Cell and Yolk Sac Hematopoietic Progenitors. Cell *109*, 29-37.

Längin, M., Mayr, T., Reichart, B., Michel, S., Buchholz, S., Guethoff, S., Dashkevich, A., Baehr, A., Egerer, S., Bauer, A.*, et al.* (2018). Consistent success in life-supporting porcine cardiac xenotransplantation. Nature *564*, 430-433.

Lee, J., Dykstra, B., Spencer, J.A., Kenney, L.L., Greiner, D.L., Shultz, L.D., Brehm, M.A., Lin, C.P., Sackstein, R., and Rossi, D.J. (2017a). mRNA-mediated glycoengineering ameliorates deficient homing of human stem cell-derived hematopoietic progenitors. J Clin Invest.

Lee, J.-H., Salci, K.R., Reid, J.C., Orlando, L., Tanasijevic, B., Shapovalova, Z., and Bhatia, M. (2017b). Brief Report: Human Acute Myeloid Leukemia Reprograming to Pluripotency Is a Rare Event and Selects for Patient Hematopoietic Cells Devoid of Leukemic Mutations. Stem cells (Dayton, Ohio).

Lee, J.H., Lee, J.B., Shapovalova, Z., Fiebig-Comyn, A., Mitchell, R.R., Laronde, S., Szabo, E., Benoit, Y.D., and Bhatia, M. (2014). Somatic transcriptome priming gates lineage-specific differentiation potential of human-induced pluripotent stem cell states. Nature communications *5*, 5605.

Lin, A., Giuliano, C., Palladino, A., John, K., Abramowicz, C., Yuan, M., Suasville, E., Lukow, D., Liu, L., Chait, A.*, et al.* (2019). Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. Science Translational Medicine *11*, eaaw8412.

Liu, G.H., Barkho, B.Z., Ruiz, S., Diep, D., Qu, J., Yang, S.L., Panopoulos, A.D., Suzuki, K., Kurian, L., Walsh, C.*, et al.* (2011). Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome. Nature *472*, 221-225.

Ljungars, A., Mårtensson, L., Mattsson, J., Kovacek, M., Sundberg, A., Tornberg, U.C., Jansson, B., Persson, N., Emruli, V.K., Ek, S.*, et al.* (2018). A platform for phenotypic discovery of therapeutic antibodies and targets applied on Chronic Lymphocytic Leukemia. npj Precision Oncology *2*, 18.

Lorenz, E., Uphoff, D., Reid, T.R., and Shelton, E. (1951). Modification of irradiation injury in mice and guinea pigs by bone marrow injections. J Natl Cancer Inst *12*, 197-201.

Lucas, D., Scheiermann, C., Chow, A., Kunisaki, Y., Bruns, I., Barrick, C., Tessarollo, L., and Frenette, P.S. (2013). Chemotherapy-induced bone marrow nerve injury impairs hematopoietic regeneration. In Nature Medicine (Nature Publishing Group), pp. 695-703.

Maciukiewicz, M., Marshe, V.S., Hauschild, A.-C., Foster, J.A., Rotzinger, S., Kennedy, J.L., Kennedy, S.H., Müller, D.J., and Geraci, J. (2018). GWAS-based machine learning approach to predict duloxetine response in major depressive disorder. Journal of Psychiatric Research *99*, 62-68.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kaminska, B., Huelsken, J., Omberg, L., Gevaert, O.*, et al.* (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell *173*, 338-354 e315.

Mancini, S.p.J.C., Mantei, N., Dumortier, A., Suter, U., MacDonald, H.R., and Radtke, F. (2005). Jagged1-dependent Notch signaling is dispensable for hematopoietic stem cell self-renewal and differentiation. Blood *105*, 2340-2342.

Martin, G. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 7634-7638.

Mathé, G., Donnall, T.E., and Ferrebee, J.W. (1959a). The Restoration Of Marrow Function After Lethal Irradiation In Man: A Review. Plastic and Reconstructive Surgery *24*.

Mathé, G., Jammet, H., Pendic, B., Schwarzenberg, L., Duplan, J.F., Maupin, B., Latarjet, R., Larrieu, M.J., Kalic, D., and Djukic, Z. (1959b). Transfusions et greffes de moelle osseuse homologus chez des humains irradies a hautes dose accidentellement Rev Fr Etudes Clin Biol *4*, 226-238.

Mauch, P. (1995). Second malignancies after curative radiation therapy for good prognosis cancers. International Journal of Radiation Oncology • Biology • Physics *33*, 959-960.

Mauch, P., Constine, L., Greenberger, J., Knospe, W., Sullivan, J., Liesveld, J.L., and Deeg, H.J. (1995). Hematopoietic stem cell compartment: Acute and late effects of radiation therapy and chemotherapy. International Journal of Radiation Oncology • Biology • Physics *31*, 1319-1339.

Maxwell, M.B., and Maher, K.E. (1992). Chemotherapy-induced myelosuppression. Seminars in Oncology Nursing *8*, 113-123.

McIntyre, B.A., Ramos-Mejia, V., Rampalli, S., Mechael, R., Lee, J.H., Alev, C., Sheng, G., and Bhatia, M. (2013). Gli3-mediated hedgehog inhibition in human pluripotent stem cells initiates and augments developmental programming of adult hematopoiesis. Blood *121*, 1543-1552.

Mikkola, H.K., and Orkin, S.H. (2006). The journey of developing hematopoietic stem cells. Development *133*, 3733-3744.

Milyavsky, M., Gan, O.I., Trottier, M., Komosa, M., Tabach, O., Notta, F., Lechman, E., Hermans, K.G., Eppert, K., Konovalova, Z.*, et al.* (2010). A distinctive DNA damage response in human hematopoietic stem cells reveals an apoptosis-independent role for p53 in self-renewal. Cell stem cell *7*, 186-197.

Mitelman, F., and Levan, G. (1981). Clustering of aberrations to specific chromosomes in human neoplasms. IV. A survey of 1,871 cases. Hereditas *95*, 79-139.

Mohrin, M., Bourke, E., Alexander, D., Warr, M.R., Barry-Holson, K., Le Beau, M.M., Morrison, C.G., and Passegue, E. (2010). Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. Cell stem cell *7*, 174-185.

Morabito, F., Mosca, L., Cutrona, G., Agnelli, L., Tuana, G., Ferracin, M., Zagatti, B., Lionetti, M., Fabris, S., Maura, F.*, et al.* (2013). Clinical monoclonal B lymphocytosis versus Rai 0 chronic lymphocytic leukemia: A comparison of cellular, cytogenetic, molecular, and clinical features. Clinical cancer research : an official journal of the American Association for Cancer Research *19*, 5890-5900.

Munoz-Novas, C., Poza-Santaella, M., Gonzalez-Gascon, Y.M.I., Hernandez-Sanchez, M., Rodriguez-Vicente, A.E., Infante, M.S., Heras, C., Foncillas, M.A., Marin, K., Hernandez-Rivas, J.M.*, et al.* (2018). The International Prognostic Index for Patients with Chronic Lymphocytic Leukemia Has the Higher Value

in Predicting Overall Outcome Compared with the Barcelona-Brno Biomarkers Only Prognostic Model and the MD Anderson Cancer Center Prognostic Index. Biomed Res Int *2018*, 9506979.

Nakanishi, M., Mitchell, R.R., Benoit, Y.D., Orlando, L., Reid, J.C., Shimada, K., Davidson, K.C., Shapovalova, Z., Collins, T.J., Nagy, A.*, et al.* (2019). Human Pluripotency Is Initiated and Preserved by a Unique Subset of Founder Cells. Cell *177*, 910-924 e922.

Nakano, H., Lyons-Cohen, M.R., Whitehead, G.S., Nakano, K., and Cook, D.N. (2017). Distinct functions of CXCR4, CCR2, and CX3CR1 direct dendritic cell precursors from the bone marrow to the lung. J Leukoc Biol *101*, 1143-1153.

Oguro, H., Ding, L., and Morrison, S.J. (2013). SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell stem cell *13*, 102-116.

Oppermann, S., Ylanko, J., Shi, Y., Hariharan, S., Oakes, C.C., Brauer, P.M., Zuniga-Pflucker, J.C., Leber, B., Spaner, D.E., and Andrews, D.W. (2016). High-content screening identifies kinase inhibitors that overcome venetoclax resistance in activated CLL cells. Blood *128*, 934-947.

Ozer, H., Armitage, J.O., Bennett, C.L., Crawford, J., Demetri, G.D., Pizzo, P.A., Schiffer, C.A., Smith, T.J., Somlo, G., Wade, J.C.*, et al.* (2000). 2000 Update of Recommendations for the Use of Hematopoietic Colony-Stimulating Factors: Evidence-Based, Clinical Practice Guidelines. Journal of Clinical Oncology *18*, 3558-3585.

Parikh, S. (2018). Chronic lymphocytic leukemia treatment algorithm 2018. Blood Cancer Journal *8*.

Parikh, S.A., Strati, P., Tsang, M., West, C.P., and Shanafelt, T.D. (2016). Should IGHV status and FISH testing be performed in all CLL patients at diagnosis? A systematic review and meta-analysis. Blood *127*, 1752-1760.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science *2*, 559-572.

Pellin, D., Loperfido, M., Baricordi, C., Wolock, S.L., Montepeloso, A., Weinberg, O.K., Biffi, A., LKlein, A.M., and Biasco, L. (2019). A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. Nature communications *10*.

Petit, I., Szyper-Kravitz, M., Nagler, A., Lahav, M., Peled, A., Habler, L., Ponomaryov, T., Taichman, R.S., Arenzana-Seisdedos, F., Fujii, N.*, et al.* (2002). G-CSF induces stem cell mobilization by decreasing bone marrow SDF-1 and up-regulating CXCR4. Nat Immunol *3*, 687-694.

Pike, B.L., and Robinson, W.A. (1970). Human bone marrow colony growth in agar-gel. Journal of Cellular Physiology *76*, 77-84.

Punzel, M., Wissink, S.D., Miller, J.S., Moore, K.A., Lemischka, I.R., and Verfaillie, C.M. (1999). The Myeloid-Lymphoid Initiating Cell (ML-IC) Assay Assesses the Fate of Multipotent Human Progenitors In Vitro. Blood *93*, 3750-3756.

Quinquenel, A., Fornecker, L.-M., Letestu, R., Ysebaert, L., Fleury, C., Lazarian, G., Dilhuydy, M.-S., Nollet, D., Guieze, R., Feugier, P.*, et al.* (2019). Prevalence of BTK and PLCG2 mutations in a real-life CLL cohort still on ibrutinib after three years: a FILO group study. Blood, blood.2019000854.

Rai, K., Sawitsky, A., Cronkite, E., Chanana, A., Levy, R., and Pasternack, B. (1975). Clinical staging of chronic lymphocytic leukemia. Blood *46*, 219-234.

Ramos-Mejia, V., Navarro-Montero, O., Ayllon, V., Bueno, C., Romero, T., Real, P.J., and Menendez, P. (2014). HOXA9 promotes hematopoietic commitment of human embryonic stem cells. Blood *124*, 3065-3075.

Ran, D., Shia, W.J., Lo, M.C., Fan, J.B., Knorr, D.A., Ferrell, P.I., Ye, Z., Yan, M., Cheng, L., Kaufman, D.S.*, et al.* (2013). RUNX1a enhances hematopoietic lineage commitment from human embryonic stem cells and inducible pluripotent stem cells. Blood *121*, 2882-2890.

Reid, J.C., Tanasijevic, B., Golubeva, D., Boyd, A.L., Porras, D.P., Collins, T.J., and Bhatia, M. (2018). CXCL12/CXCR4 Signaling Enhances Human PSC-Derived Hematopoietic Progenitor Function and Overcomes Early In Vivo Transplantation Failure. Stem Cell Reports *10*, 1625-1641.

Richman, C.M., Weiner, R.S., and Yankee, R.A. (1976). Increase in circulating stem cells following chemotherapy in man. Blood *47*, 1031-1039.

Roberts, A.W., Davids, M.S., Pagel, J.M., Kahl, B.S., Puvvada, S.D., Gerecitano, J.F., Kipps, T.J., Anderson, M.A., Brown, J.R., Gressick, L.*, et al.* (2016). Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. N Engl J Med *374*, 311-322.

Rodriguez-Lara, V., Ignacio, G.S., and Cerbon Cervantes, M.A. (2017). Estrogen induces CXCR4 overexpression and CXCR4/CXL12 pathway activation in lung adenocarcinoma cells in vitro. Endocr Res, 1-13.

Rossi, D., Spina, V., and Gaidano, G. (2018). Biology and treatment of Richter syndrome. Blood *131*, 2761-2772.

Sachlos, E., Risueno, R.M., Laronde, S., Shapovalova, Z., Lee, J.H., Russell, J., Malig, M., McNicol, J.D., Fiebig-Comyn, A., Graham, M.*, et al.* (2012). Identification of drugs including a dopamine receptor antagonist that selectively target cancer stem cells. Cell *149*, 1284-1297.

Saxena, S., Ronn, R.E., Guibentif, C., Moraghebi, R., and Woods, N.B. (2016). Cyclic AMP Signaling through Epac Axis Modulates Human Hemogenic Endothelium and Enhances Hematopoietic Cell Generation. Stem Cell Reports.

Silberstein, L., Goncalves, K.A., Kharchenko, P.V., Turcotte, R., Kfoury, Y., Mercier, F., Baryawno, N., Severe, N., Bachand, J., Spencer, J.A.*, et al.* (2016). Proximity-Based Differential Single-Cell Analysis of the Niche to Identify Stem/Progenitor Cell Regulators. Cell stem cell.

Sokolov, A., Carlin, D.E., Paull, E.O., Baertsch, R., and Stuart, J.M. (2016a). Pathway-Based Genomics Prediction using Generalized Elastic Net. PLoS Comput Biol *12*, e1004790.

Sokolov, A., Paull, E.O., and Stuart, J.M. (2016b). One-class detection of cell states in tumor subtypes. Pac Symp Biocomput *21*, 405-416.

Souers, A.J., Leverson, J.D., Boghaert, E.R., Ackler, S.L., Catron, N.D., Chen, J., Dayton, B.D., Ding, H., Enschede, S.H., Fairbrother, W.J.*, et al.* (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nat Med *19*, 202-208.

Spaner, D.E., Lee, E., Shi, Y., Wen, F., Li, Y., Tung, S., McCaw, L., Wong, K., Gary-Gouy, H., Dalloul, A.*, et al.* (2013). PPAR-alpha is a therapeutic target for chronic lymphocytic leukemia. Leukemia *27*, 1090-1099.

Stacchini, A., Aragno, M., Vallario, A., Alfarano, A., Circosta, P., Gottardi, D., Faldella, A., Rege-Cambrin, G., Thunberg, U., Nilsson, K.*, et al.* (1999). MEC1 and MEC2: two new cell lines derived from B-chronic lymphocytic leukaemia in prolymphocytoid transformation. Leukemia Research *23*, 127-136.

Stevens, L.C., and Little, C.C. (1954). Spontaneous Testicular Teratomas in an Inbred Strain of Mice. Proceedings of the National Academy of Sciences of the United States of America *40*, 1080-1087.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e1821.

Sturgeon, C.M., Ditadi, A., Awong, G., Kennedy, M., and Keller, G. (2014). Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. Nature biotechnology *32*, 554-561.

Sturgeon, C.M., Ditadi, A., Clarke, R.L., and Keller, G. (2013). Defining the path to hematopoietic stem cells. Nature biotechnology *31*, 416-418.

Subramaniam, A., Safaee Talkhoncheh, M., Magnusson, M., and Larsson, J. (2018). Endothelial protein C receptor (EPCR) expression marks human fetal liver hematopoietic stem cells. Haematologica.

Sugimura, R., Jha, D.K., Han, A., Soria-Valles, C., da Rocha, E.L., Lu, Y.-F., Goettel, J.A., Serrao, E., Rowe, R.G., Malleshaiah, M.*, et al.* (2017). Haematopoietic stem and progenitor cells from human pluripotent stem cells. Nature.

Suzuki, N., Yamazaki, S., Yamaguchi, T., Okabe, M., Masaki, H., Takaki, S., Otsu, M., and Nakauchi, H. (2013). Generation of Engraftable Hematopoietic Stem Cells From Induced Pluripotent Stem Cells by Way of Teratoma Formation. Molecular Therapy *21*, 1424-1431.

Tachibana, C. (2015). Transcriptomics today: Microarrays, RNA-seq, and more. Science.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861-872.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663-676.

Talib, S., and Shepard, K.A. (2020). Unleashing the cure: Overcoming persistent obstacles in the translation and expanded use of hematopoietic stem cell-based therapies. Stem cells translational medicine.

The International Stem Cell Initiative, Adewumi, O., Aflatoonian, B., Ährlund-Richter, L., Amit, M., Andrews, P.W., Beighton, G., Bello, P.A., Benvenisty, N., Berry, L.S*., et al.* (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nature biotechnology *25*, 803-816.

The International Stem Cell Initiative, Amps, K., Andrews, P.W., Anyfantis, G., Armstrong, L., Avery, S., Baharvand, H., Baker, J., Baker, D., Munoz, M.B*., et al.* (2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nature biotechnology *29*, 1132-1144.

The World Health Organization, Stewart, B.W., and Wild, C.P. (2014). World Cancer Report 2014. Chapter 5.13.

Thomas, E.D., Lochte, H.L., Cannon, J.H., Sahler, O.D., and Ferrebee, J.W. (1959). Supralethal whole body irradiation and isologous marrow transplantation in man. J Clin Invest *38*, 1709-1716.

Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., Becker, R.A., and Hearn, J.P. (1995). Isolation of a primate embryonic stem cell line. Proceedings of the National Academy of Sciences *92*, 7844-7848.

Thomson, J.A.I.-E., J.; Shapiro, S. S.; Waknitz, M. A.; Swiergiel, J. J.; Marshall, V. S.; Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. Science (New York, NY) *6*, 1145-1147.

Tikhonova, A.N., Dolgalev, I., Hu, H., Sivaraj, K.K., Hoxha, E., Cuesta-Domínguez, Á., Pinho, S., Akhmetzyanova, I., Gao, J., Witkowski, M*., et al.* (2019). The bone marrow microenvironment at single-cell resolution. Nature.

Tracy, S., Yuan, G.-C., and Dries, R. (2019). RESCUE: imputing dropout events in single-cell RNA-sequencing data. BMC Bioinformatics *20*, 388.

Vahdat, L.T., Balmaceda, C., Papadopoulos, K., Frederick, D., Donovan, D., Sharpe, E., Kaufman, E., Savage, D., Tiersten, A., Nichols, G*., et al.* (2002). Phase II trial of sequential high-dose chemotherapy with paclitaxel, melphalan and cyclophosphamide, thiotepa and carboplatin with peripheral blood progenitor support in women with responding metastatic breast cancer. Bone Marrow Transplant *30*, 149-155.

van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of machine learning research : JMLR *9*, 2579-2605.

van der Straten, L., Levin, M.D., Visser, O., Posthuma, E.F.M., Doorduijn, J.K., Kater, A.P., and Dinmohamed, A.G. (2020). Survival continues to increase in chronic lymphocytic leukaemia: a population-based analysis among 20 468 patients diagnosed in the Netherlands between 1989 and 2016. Br J Haematol.

Velardi, E., Tsai, J.J., Radtke, S., Cooper, K., Argyropoulos, K.V., Jae-Hung, S., Young, L.F., Lazrak, A., Smith, O.M., Lieberman, S*., et al.* (2018). Suppression of luteinizing hormone enhances HSC recovery after hematopoietic injury. Nature Medicine *24*, 239-246.

Vo, L.T., Kinney, M.A., Liu, X., Zhang, Y., Barragan, J., Sousa, P.M., Jha, D.K., Han, A., Cesana, M., Shao, Z*., et al.* (2018). Regulation of embryonic haematopoietic multipotency by EZH1. Nature.

Vodyanik, M.A., Bork, J.A., Thomson, J.A., and Slukvin, I. (2005). Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. Blood *105*, 617-626.

Wang, J.C.Y., Doedens, M., and Dick, J.E. (1997). Primitive Human Hematopoietic Cells Are Enriched in Cord Blood Compared With Adult Bone Marrow or Mobilized Peripheral Blood as Measured by the Quantitative In Vivo SCID-Repopulating Cell Assay. Blood *89*, 3919-3924.

Wang, L., Li, L., Shojaei, F., Levac, K., Cerdan, C., Menendez, P., Martin, T., Rouleau, A., and Bhatia, M. (2004). Endothelial and hematopoietic cell fate of human embryonic stem cells originates from primitive endothelium with hemangioblastic properties. Immunity *21*, 31-41.

Wang, L., Menendez, P., Shojaei, F., Li, L., Mazurier, F., Dick, J.E., Cerdan, C., Levac, K., and Bhatia, M. (2005). Generation of hematopoietic repopulating cells from human embryonic stem cells independent of ectopic HOXB4 expression. The Journal of experimental medicine *201*, 1603-1614.

Wang, L.M., P.; Shojaei, F.; Li, L.; Mazurier, F.; Dick, J. E.; Cerdan, C.; Levac, K.; Bhatia, M. (2005). Generation of hematopoietic repopulating cells from human embryonic stem cells independent of ectopic HOXB4 expression. Journal of Experimental Medicine *201*, 1603-1614.

Wang, Z.-Y., and Chen, Z. (2008). Acute promyelocytic leukemia: from highly fatal to highly curable. Blood *111*, 2505-2515.

Wendel-Hansen, V., Sällström, J., De Campos-Lima, P.O., Kjellström, G., Sandlund, A., Siegbahn, A., Carlsson, M., Nilsson, K., and Rosén, A. (1994). Epstein-Barr virus (EBV) can immortalize B-cll cells activated by cytokines. Leukemia *8*, 476-484.

Werbowetski-Ogilvie, T.E., Bosse, M., Stewart, M., Schnerch, A., Ramos-Mejia, V., Rouleau, A., Wynder, T., Smith, M.J., Dingwall, S., Carter, T.*, et al.* (2009). Characterization of human embryonic stem cells with features of neoplastic progression. Nature biotechnology *27*, 91-97.

Woyach, J.A. (2019). Can Combination Targeted Therapy Bring About Another Paradigm Shift in Chronic Lymphocytic Leukemia? Journal of Clinical Oncology, JCO.19.01718.

Wu, A.M., Till, J.E., Siminovitch, L., and McCulloch, E.A. (1968). Cytological Evidence For A Relationship Between Normal Hematopoietic Colony-Forming Cells And Cells Of The Lymphoid System. The Journal of experimental medicine *127*, 455-464.

Wu, J., Platero-Luengo, A., Sakurai, M., Sugawara, A., Gil, M.A., Yamauchi, T., Suzuki, K., Bogliotti, Y.S., Cuello, C., Morales Valencia, M.*, et al.* (2017). Interspecies Chimerism with Mammalian Pluripotent Stem Cells. Cell *168*, 473-486 e415.

Xu, M.-j., Matsuoka, S., Yang, F.-C., Ebihara, Y., Manabe, A., Tanaka, R., Eguchi, M., Asano, S., Nakahata, T., and Tsuji, K. (2001). Evidence for the presence of murine primitive megakarycytopoiesis in the early yolk sac. Blood *97*, 2016-2022.

Yang, M., McKay, D., Pollard, J.W., and Lewis, C.E. (2018). Diverse Functions of Macrophages in Different Tumor Microenvironments. Cancer Res *78*, 5492-5503.

Yin, S., Gambe, R.G., Sun, J., Martinez, A.Z., Cartun, Z.J., Regis, F.F.D., Wan, Y., Fan, J., Brooks, A.N., Herman, S.E.M.*, et al.* (2019). A Murine Model of Chronic Lymphocytic Leukemia Based on B Cell-Restricted Expression of Sf3b1 Mutation and Atm Deletion. Cancer Cell *35*, 283-296 e285.

Zenz, T., Eichhorst, B., Busch, R., Denzel, T., Habe, S., Winkler, D., Buhler, A., Edelmann, J., Bergmann, M., Hopfinger, G.*, et al.* (2010). TP53 mutation and survival in chronic lymphocytic leukemia. J Clin Oncol *28*, 4473-4479.

Zhang, Y., Depond, M., He, L., Foudi, A., Kwarteng, E.O., Lauret, E., Plo, I., Desterke, C., Dessen, P., Fujii, N.*, et al.* (2016). CXCR4/CXCL12 axis counteracts hematopoietic stem cell exhaustion through selective protection against oxidative stress. Sci Rep *6*, 37827.

Zheng, G., Terry, J., Belgrader, P., Ryvkin, P., Bent, Z., Wilson, R., Ziraldo, S., Wheeler, T., McDermott, G., Zhu, J.*, et al.* (2017). Massively parallel digital transcriptional profiling of single cells. Nature communications *8*, 1-12.

Zinzani, P.L., Rambaldi, A., Gaidano, G., Girmenia, C., Marchetti, M., Pane, F., Tura, S., and Barosi, G. (2019). Infection control in patients treated for chronic lymphocytic leukemia with ibrutinib or idelalisib: recommendations from Italian society of hematology. Leuk Res *81*, 88-94.

Appendix II: Copyrights and permissions to reprint published material

Chapter 3: CXCL12/CXCR4 signaling enhances human PSC-derived hematopoietic progenitor function and overcomes early in vivo transplantation failure
Reid JC, Tanasijevic B, Golubeva D, Boyd AL, Porras DP, Collins TJ, Bhatia M.
Stem Cell Reports Article: Volume 10, Issue 5, pg. 1625-1641, May 08, 2018
DOI: https://doi.org/10.1016/j.stemcr.2018.04.003

Stem Cell Reports is an open access journal which is published for the International Society for Stem Cell Research (ISSCR) by *Cell Press*. Authors are asked to sign a nonexclusive publishing agreement that allows them to retain copyright of their work.

Appendix III: List of Scientific Publications

Published refereed papers

Nakanishi M, Mitchell MR, Benoit YD, Orlando L, **Reid JC**, Shimada K, Davidson KC, Shapovalova Z, Collins TJ, Nagy A, Bhatia M. Human pluripotency is initiated and preserved by a unique subset of founder cells. ***Cell*** 177(4): p910-924 (2019). PMID: 30982595.

Reid JC, Tanasijevic B*, Golubeva D*, Boyd AL, Porras DP, Collins TJ, Bhatia M. CXCL12/CXCR4 Signaling Enhances Human PSC-Derived Hematopoietic Progenitor Function and Overcomes Early *In Vivo* Transplantation Failure. *Stem Cell Reports* 10(5): p1625-1641 (2018). PMID: 29742393.

Boyd AL*, Aslostovar L*, **Reid JC**, Ye W, Tanasijevic B, Porras DP, Shapovalova Z, Almakadi M, Foley R, Leber B, Xenocostas A, Bhatia M. Identification of Chemotherapy-Induced Leukemic-Regenerating Cells Reveals a Transient Vulnerability of Human AML Recurrence. ***Cancer Cell*** 34(3): p483-498 (2018). PMID: 30205048.

Boyd AL, **Reid JC**, Salci KR, Aslostovar L, Benoit YD, Shapovalova Z, Nakanishi M, Porras DP, Almakadi M, Campbell CJV, Jackson MF, Ross CA, Foley R, Leber B, Allan DS, Sabloff M, Xenocostas A, Collins TJ, Bhatia M. Acute myeloid leukaemia disrupts endogenous myelo-erythropoiesis by compromising the adipocyte bone marrow niche. ***Nature Cell Biology*** 19(11): p1336-1347 (2017). PMID: 29035359.

Lee JH*, Salci KR*, **Reid JC**, Orlando L, Tanasijevic B, Shapovalova Z, Bhatia M. Pluripotent reprogramming selects for AML patient hematopoietic cells devoid of leukemic mutations. *Stem Cells* 35(9): p2095-2102 (2017). PMID: 28758276.

Manuscripts submitted to peer-reviewed journals

Reid JC, Golubeva D, Orlando L, Boyd AL, Leber A, Hébert J, Gentile M, Agnelli L, Morabito F, Neri A, Leber B, Campbell CJV, Bhatia, M. Molecular targets of trisomy 12 in chronic lymphocytic leukemia uncovered using human pluripotency. ***Science Translational Medicine*** (2019).

Orlando L*, Benoit YD*, Nakanishi M, **Reid JC**, Boyd AL, Tanasjievic B, Restall IJ, Aslostovar L, Lu JD, Ye W, Laronde S, Doyle MS, Shapovalova Z, García-Rodriguez JL, Bergin CJ, Masibag AN, Luchman HA, Collins TJ, Weiss S, Bhatia M. Targetable programs of human cancer are ontogenically rooted within cell fate trajectories of pluripotency. ***Nature*** (2020).

Manuscripts awaiting submission

Orlando L*, Tanasijevic B*, **Reid JC**, Nakanishi M, García-Rodriguez JL, Kapil C, Aslostovar L, Lu JD, Shapovalova Z, Mahammad S, Mitchell RR, Mickie Bhatia M. Histone variant H2A.X phosphorylation regulates cell fate of human stem cells. *In preparation for Cell Stem Cell*.

Aslostovar L, Boyd AL, Benoit YD, Lu JD, Nakanishi M, Porras DP, **Reid JC**, Leber B, Xenocostas A, Roley R, Bhatia M. Acquisition of dopamine receptor mediated signaling allows selective therapeutic targeting of neoplastic progenitors in AML patients. *In preparation for Stem Cells*.

*(*These authors contributed equally)*

Appendix IV: Ethics Approvals

McMaster University

ANIMAL Utilization Protocol

AREB Office Use Only

Approval Date: 2 - Dec - 2019 APPROVED AUP # 19 - 11 - 29

Expiry Date: 2 - Dec - 2023 Replaces AUP # 16 - 10 - 37

University Veterinarian's Signature — 2 Dec 2019 (Date) — AREB Chair's Signature — 2 Dec 2019 (Date)

Principal Investigator's Signature — Oct 7/2019 (Date)

Prepared By:
Name: Emily Broder
Email Address: brodere@mcmaster.ca

Refer to the Guide for Preparation of AUPs for assistance with completing this form
(available at http://www.fhs.mcmaster.ca/healthresearch/areb_forms.html)

Section 1 *Project Title*

Title: Engraftment of immune-deficient mice using murine and human stem cells.

Type of Project (Check all applicable): ☒ Research ☐ Teaching* ☐ New Project ☒ Ongoing Project ☐ Research Pilot Study ☐ Other

*Attach Teaching Addendum available at http://www.fhs.mcmaster.ca/healthresearch/areb_forms.html.

Section 2 *Principal Investigator*

Last Name	Bhatia	First Name	Mickie
Title	Professor	Department	Stem Cell and Cancer Research Institute (SCC-RI)
Business Phone	905-525-9140 x 28687	Home Phone	905-537-2166
Emergency Phone	905-537-2166	Cell/Pager #	N/A
Laboratory Room #	MDCL 5116	Laboratory Phone	905-525-9140 x 21872
Institutional Email	mbhatia@mcmaster.ca		
Mailing Address	MDCL 5029		

Section 3 *Personnel and Training*

List all staff personnel working under this AUP. Provide at least one person's after hours emergency contact number.

Name	Department	Title Co-PI, Tech, Undergrad Student, Grad	Extension	After Hours Emergency Contact Number
Allison Boyd	SCC-RI	Post-Doc	21872	289-244-8010
Jennifer Reid	SCC-RI	Grad Student	21872	289-439-4753
Juan Luis Garcia-Rodriguez	SCC-RI	Post-Doc	21872	905-923-2824
SEE ANNEX 1 FOR MORE CONTACT				

Is this work being conducted in collaboration with any other Institutions / Facilities? ☐ Yes ☒ No

Please name the institution / facility.(Refer to the AREB Institutional Collaboration Policy (https://fhs.mcmaster.ca/healthresearch/documents/InstitutionalCollaborationsPolicy_000.pdf)

If yes, please provide details of the collaboration (work being carried out, what facility the work in taking place at, who is performing the procedures)

McMaster University

Biohazard Utilization Protocol

McMaster Biosafety Office
Health Science Centre 1J11A
1200 Main Street West
Hamilton, Ontario, CANADA
L8N 3Z5

Phone 905.525.9140
Ext 27527
Fax 905.528.8539
Email:hainesk@mcmaster.ca
https://biosafety.mcmaster.ca

This BUP represents the current information the Biosafety Office has on file as of 2020-01-15. Please ensure all listed personnel have access to this document. For proof of project-specific compliance, please email hainesk@mcmaster.ca. For all other external requests for proof of biohazard approval, please contact robertiv@mcmaster.ca for the appropriate documentation. This document contains protected information, do not send it to persons outside of McMaster University.

Name:	**Bhatia , Mick**	Department:	Biochemistry & Biomedical Sciences	BUPID:	93
Bldg/Room:	MDCL 5029	Phone:	28687	Emergency Phone#:	--private info--
Email:	mbhatia@mcmaster.ca	Requested CL:	Level 2 enhanced	Contact Person:	Tony Collins
BUP Status:	Approved	Renewal:	2022-06-10	Contact Email:	tjc@mcmaster.ca

Applicable Containment Standards:
- CL1 - Canadian Biosafety Guideline (also refer to BMBL (NIH/CDC) and WHO Laboratory Biosafety Manual (WHO))
- CL2 - Canadian Biosafety Standard

Scope: Characterization of molecular pathways regulating human hematopoietic and pluripotent stem cells including Notch, Sonic Hedgehog, Bone Morphogenetic Proteins, and human Wingless homologs involved in Wnt signaling transduction. These pathways are linked to the uncontrolled growth of tissue initiated by purified stem cells and are important to our understanding of early events in human cellular transformation. Identification of target genes regulated by mesodermal factors and genes involved in controlling self-renewal and differentiation of primary human blood stem cells based on differential expression and ontogenic mechanisms. Creation of novel in vivo models for cellular/tissue regeneration through transplantation of human stem cells.

The following persons have online access to this BUP. For information on workers, see section 6 below.

Name
Bhatia, Mick
Collins, Tony

Rooms with (CF) are designated as central facility rooms and will be audited under the facility director's BUP.

Building	Room	Rad	Animal	Dedicated Waste Room	Room Type	Extension	CL	CO2 Cyl	O2 Cyl	O2 CO2 Mix Cyl	N2 Cyl	He2 Cyl	Ar Cyl	Air Cyl	+4 C	-20 C	-80 C	-150 C	LN2	Dry Inc	CO2 Inc
HSC	1U02M				Animal Postmortem Room		Level 2														
HSC	1U02N				Animal Housing Room		Level 1														
HSC	1U02P				Animal Housing Room		Level 2														
HSC	1U02Q				Animal Procedure Room		Level 2														
HSC	1U02W				Animal Procedure Room		Level 2														
MDCL	5059				Laboratory		Level 1														
MDCL	5060				Equipment Room		Level 1														
MDCL	5072				Waste Room		Waste Room														
MDCL	5075A				Laboratory		Level 1														
MDCL	5075B				Equipment Room		Level 1														
MDCL	5075C				Laboratory		Level 1														
MDCL	5076				Laboratory	21010	Level 1														
MDCL	5119 (Tissue Culture Ante Room)				Corridor		Level 1														
MDCL	5120				Storage Room	26252															
MDCL	5124				Laboratory		Level 2														
MDCL	5125				Laboratory		Level 2 enhanced														
MDCL	5127				Laboratory	26252	Level 1														
MDCL	5142				Laboratory		Level 2														
MDCL	5143				Equipment Room		Level 2														
MDCL	5144 (CF)				Laboratory		Level 2														

RECEIVED

HiREB
Hamilton Integrated Research Ethics Board

Hamilton Integrated Research Ethics Board
RENEWAL FORM

REB Project #: 08-330

Principal Investigator: Dr. Mickie Bhatia

Project Title: Characterisation of Human Blood Stem Cells

[X] **Approved for Continuation**

[ ] **Approved conditional on changes noted below**

Type of Approval:

[X] Full Research Ethics Board
[ ] Research Ethics Board Executive

REB Approval Period: **Approval period covers July 15-2019 to July 15-2020**

The Hamilton Integrated Research Ethics Board operates in compliance with and is constituted in accordance with the requirements of: The Tri-Council Policy Statement on Ethical Conduct of Research Involving Humans; The International Conference on Harmonization of Good Clinical Practices; Part C Division 5 of the Food and Drug Regulations of Health Canada, and the provisions of the Ontario Personal Health Information Protection Act 2004 and its applicable Regulations; for studies conducted at St. Joseph's Hospital, HIREB complies with the health ethics guide of the Catholic Alliance of Canada.

Frederick A. Spencer, MD, Chair
Mark Inman, MD, PhD, Chair

7/3/2019
Date of REB Meeting

All Correspondence should be addressed to the HIREB Chair(s) and forwarded to:
HIREB Coordinator
293 Wellington St. N, Suite 102, Hamilton ON L8L 8E7
Tel. 905-521-2100 Ext. 42013 Fax: 905-577-8378

HiREB
Hamilton Integrated Research Ethics Board

Hamilton Integrated Research Ethics Board
RENEWAL FORM

REB Project #: 08-042-T

Principal Investigator: Dr. Michael Trus

Project Title: Stem Cell and Cancer Research Institute - Live Cell Bank

[X] **Approved for Continuation**

[ ] **Approved conditional on changes noted below**

Type of Approval:

[X] Full Research Ethics Board
[ ] Research Ethics Board Executive

REB Approval Period: **Approval period covers January 17-2019 to January 17-2020**

[ ] **New Enrolment Suspended**

[ ] **Suspended pending further review**

The Hamilton Integrated Research Ethics Board operates in compliance with and is constituted in accordance with the requirements of: The Tri-Council Policy Statement on Ethical Conduct of Research Involving Humans; The International Conference on Harmonization of Good Clinical Practices; Part C Division 5 of the Food and Drug Regulations of Health Canada, and the provisions of the Ontario Personal Health Information Protection Act 2004 and its applicable Regulations; for studies conducted at St. Joseph's Hospital, HIREB complies with the health ethics guide of the Catholic Alliance of Canada.

Frederick A. Spencer, MD, Chair
Mark Inman, MD, PhD, Chair

2/19/2019
Date of REB Meeting

All Correspondence should be addressed to the HIREB Chair(s) and forwarded to:
HIREB Coordinator
293 Wellington St. N, Suite 102, Hamilton ON L8L 8E7
Tel. 905-521-2100 Ext. 42013 Fax: 905-577-8378

Canadian Institutes of Health Research
160 Elgin Street, 9th Floor
Address Locator 4809A
Ottawa, Ontario K1A 0W9

Instituts de recherche en santé du Canada
160, rue Elgin, 9e étage
Indice de l'adresse 4809A
Ottawa (Ontario) K1A 0W9

August 16, 2018

Dr. Mickie Bhatia
Stem Cell and Cancer Research Institute (SCC-RI)
McMaster University, Michael G. DeGroote Centre for Learning
1280 Main Street West, MDCL 5029
Hamilton, ON L8S 4K1

Re: Stem Cell Oversight Committee review of CIHR Foundation Grant, entitled "Molecular Targeting of Human Acute Myeloid Leukemia and Bone Marrow Regeneration" (application #:388945)

Dear Dr. Bhatia:

The Stem Cell Oversight Committee (SCOC) met by teleconference on August 8, 2018 to review your stem cell research application for conformity to the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*, TCPS 2 (2014), chapter 12, section F.

The Committee has approved your request to use the SCOC-approved H1, H9, and CA2 hESC lines and human iPS cells. Since the proposed research falls under Governing Council's delegation of authority, this letter constitutes final approval by CIHR of the above-mentioned project. As CIHR does not notify co-applicants, we ask that you please inform those individuals and their research institutions (if different from your own) of the outcome of this application.

If you wish to conduct experiments involving human pluripotent stem cells that are not described in this application, you will need to seek SCOC approval by submitting a description of the proposed human pluripotent stem cell research. SCOC will send you a reminder of this requirement on an annual basis.

Should you wish to use additional SCOC-approved hESC lines not described in the original SCOC application for this research project, you need only to notify SCOC in writing. You should include the exact title of your original application and date of submission. There will be no need to submit an amended application for SCOC review.

SCOC should also be advised of any conflicts of interest that may arise during the course of this research.

All the best in your research endeavours.

Sincerely yours,

for

Anne Martin-Matthews, O.C., Ph.D., FCAHS
Acting Vice-President, Research, Knowledge Translation and Ethics

cc: Bernard Thébaud, Chair, Stem Cell Oversight Committee
Marta Arnaldo, Manager, Governance Secretariat
Michelle Peel, A/Director General, Science Policy Branch