Complex alternative splicing patterns in human hematopoietic cell subpopulations revealed by third-generation long reads

A. Deslattes Mays[1], E. Tseng[2], R. Sebra[3], P. Baybayan[2], M. Schmidt[1], A. Wellstein[1]

[1]Department of Oncology and Medicine, Lombardi Cancer Center, Washington DC 20057 [2]Pacific Biosciences, Menlo Park, CA 94025

[3]Icahn School of Medicine at Mount Sinai, Institute for Genomics and Multiscale Biology, New York, NY 10029

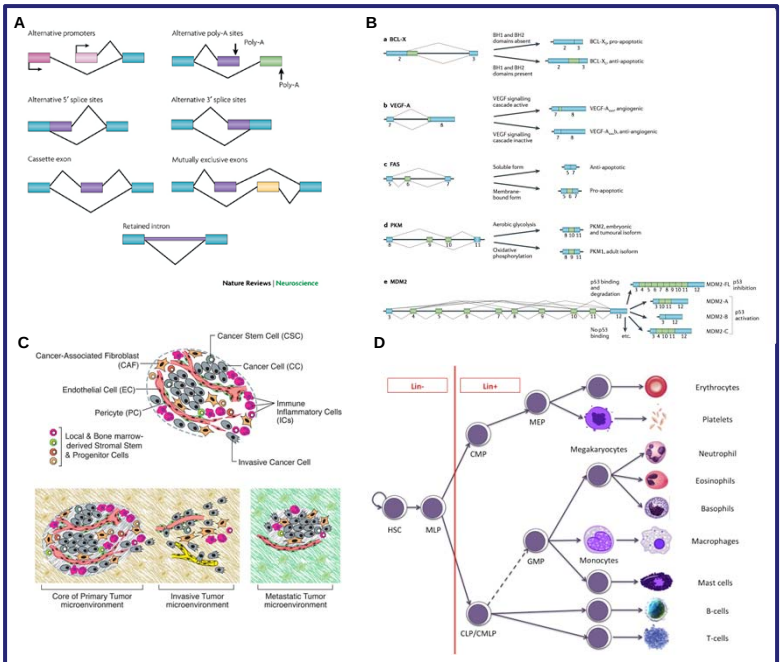

***Figure 1* – Alternative splicing, cancer progression and *hematopoetic cell differentiation*- A.** Alternative splicing discovered in the late 1970s[1]. Verified experimentally in the 1980s[2]. Real revolution came with the sequencing of the human genome[3]. Before the sequencing of the human genome – estimates of gene counts in the human genome ranged from 20-100,000. Alternative splicing provides multiple transcript isoforms for the same transcript parts[2]. **B**. Examples of cases where opposite function from a genes transcribed from the same genomic locus results in opposite function. **C.** Probing bone marrow hematopoiesis provides insight into cancer stem cells and their differentiating programs[3]. Isolating the most undifferentiated cells from a healthy bone marrow cell population and measuring the structure and the abundance helps us to understand health which gives us insight into disease. **D.** Hematopoetic cell lineage differentiation, focusing on the early upstream uncommitted cells.

[1]Gilbert, W.. "Why genes in pieces? Nature 271, no 5645 (1978): 501

[2]Florea, L. "Bioinformatics of alternative splicing and its regulation." Briefings in Bioinfo. 7.1 (2006): 55-69.

[3]Bonnal, Sophie, Luisa Vigevani, and Juan Valcárcel. "The spliceosome as a target of novel antitumour drugs." Nature Reviews Drug Discovery 11.11 (2012): 847-859.

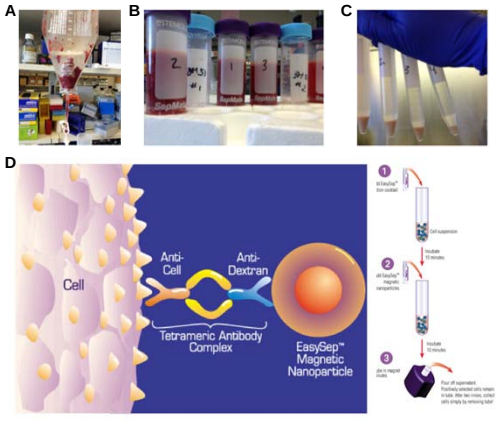

***Figure 2 – Lineage negative cells isolated through magnetic separation.* A.** Discarded collection bone marrow bags obtained from Georgetown MedStar's cell processing unit. **B**. Bone marrow transferred to SepMate™- 50 ml tubes layered with Ficol and centrifuged obtaining bone marrow mononuclear cells (BMNCs). **C.** BMNCs purified with ammonium chloride. **D.** Lineage negative (Lin-) isolated through magnetic separation using an antibody cocktail in tetrameric complex with EasySep™ magnetic nanoparticles. Resulting in 1-2% lineage negative (Lin-) cells separated from 98-99% lineage positive (Lin+) cells from initial total BMNC cell population.

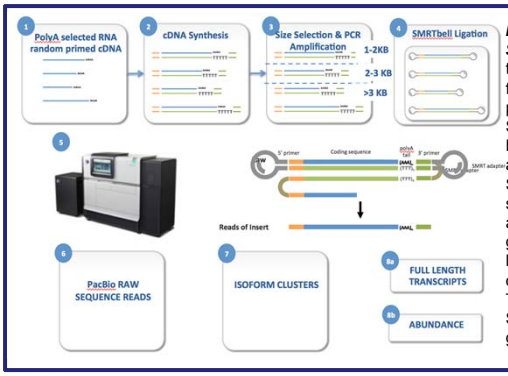

Figure 3 - Log read RNA-Sequencing Workflow. PacBio transcript sequencing (Iso-Seq™), full-length cDNA synthesis from polyA RNA using the Clontech® SMARTer™ PCR cDNA Synthesis Kit; size selection of 1-2 kb, 2-3 kb, and 3-6 kb fractions: conversion to SMRTbell™ libraries. BluePippin™ size selection protocol includes 2 amplification steps, before and after gel sizing. Full-length cDNA libraries can be produced from 1 ng of poly(A) RNA, or 2 ng total RNA. The total time from input RNA to SMRTbell template library generation is 3 to 4 days

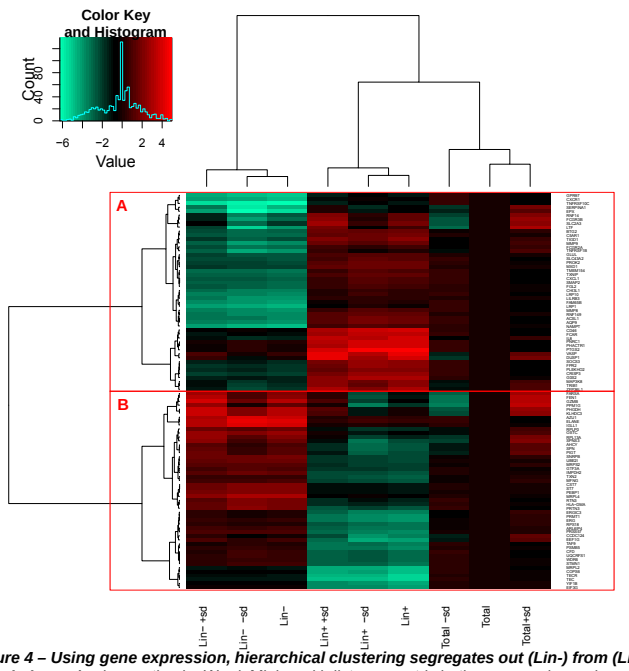

Figure 4 – Using gene expression, hierarchical clustering segregates out (Lin-) from (Lin+) cell populations. Agglomeration by Ward, Minkowski distance metric both genes and samples segregate (Lin-) and (Lin+) cell populations. Normalized by the total cell population, the top 50 genes sorted by expression clearly segregate the into two groups. A represents those genes enriched in the (Lin+) cell population and B represents those cells enriched in the (Lin-) cell populations.

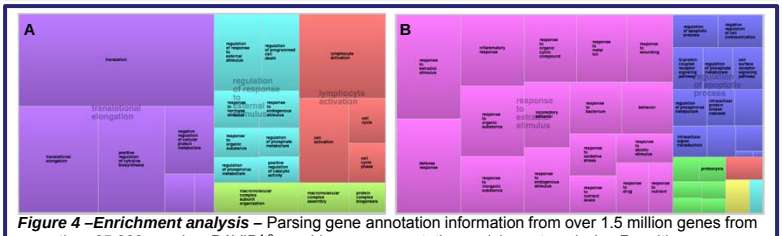

Figure 4 –Enrichment analysis – Parsing gene annotation information from over 1.5 million genes from more than 65,000 species, DAVID[1,2] provides gene-annotation enrichment analysis. Resulting gene ontologies and their associated p-values are provided to REVIGO[3] for further enrichment analysis. **A.** (Lin+) biological processes from top genes expressed relative to total and differentially expressed versus (Lin-). **B.** (Lin-) biological processes from top gene expressed relative to total and differentially expressed versus (Lin-).[1] Huang DW, Sherman BT, Lempicki RA. Nature Protoc. 2009;4(1):44-57. [2]Huang DW, Sherman BT, Lempicki RA. Nucleic Acids Res. 2009;37(1):1-13. [3]Supek F, Bošnjak M, Škunca N, Šmuc T. - PLoS ONE 2011. doi:10.1371/journal.pone.0021800

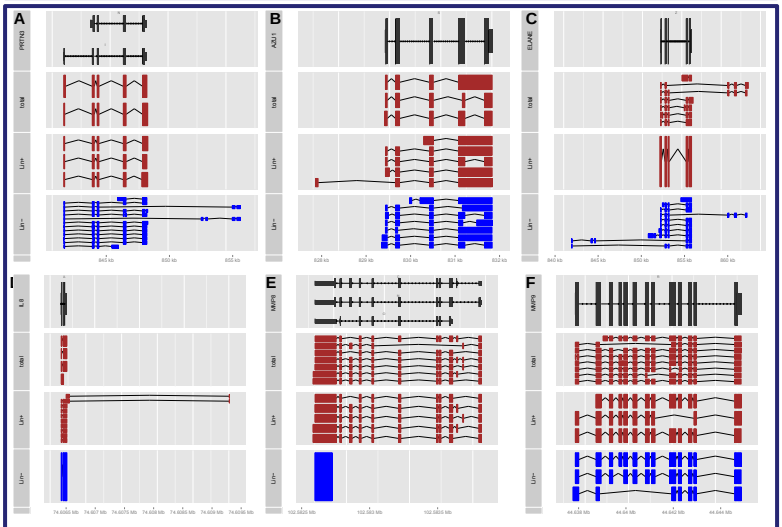

***Figure 6 – Complex patterns transcript isoforms shown in (Lin-) and (Lin+) revealed through third generation sequencing.* A.** PRTN3, **B.** AZU1 and **C.** ELANE genes highly expressed in the (Lin-) cell population are co-linear on the genome. AZU1, ELANE, PRTN3 and the CFD are often co-expressed and are under evolutionary pressure to remain heterozygous. **D.** IL8 , **E.** MMP8 and **F.** MMP9 genes highly expressed in (Lin+) cell population. The metal proteases, in particular MMP8 have been shown to play a role in IL-8 induced mobilization of hematopoietic progenitor cells from the bone marrow.

PACIFIC BIOSCIENCES™

Lombardi COMPREHENSIVE CANCER CENTER

GEORGETOWN UNIVERSITY