Predictive value genes in AML

Prognostic tools in AML

The **cytogenetic karyotype is the most important prognostic factor** in acute myeloid leukemia (AML), it's the basis of the current risk classification in the disease (Grimwade, Hematol Oncol Clin North Am 2011).

<Normal cytogenetic karyotype = no structural variations in chromosomes (big chromosomal aberrations ≠ mutations)>

However, **40–45% of all AML patients are cytogenetic normal** (CN) (Grimwade, Blood 2010; Lazarevic, Blood Cancer J 2014).

CN-AML can still have mutations.

Driver mutations in AML

- **AML genomes have fewer mutations than most other adult cancers**, with an average of only 13 mutations found in genes (TCGA Research Network, N Engl J Med 2013)
- genetic abnormalities in leukemia can be roughly grouped into 2 classes: (1) mutations involving signal transduction pathways and giving rise to proliferative advantages to leukemia clones (2) those affecting transcription factors or cofactors and causing impaired differentiation (Schlenk et al., N Engl J Med 2008)
- Shen et al., Blood 2011: **74.7% samples (N=605) have at least 1 mutation**

Prognostic panels (mut) for risk stratification in AML

- European Leukemia Net (ELN) risk classification system (cytogenetic)
- **18 genes and their somatic mutations** for subgroups for disease prognosis by Patel et al., N Engl J Med 2012
- molecular-based classification system based on **mutation panel of 76 genes** by Papaemmanuil et al., N Engl J Med 2016
- TCGA Research Network, N Engl J Med 2013: **23 genes significantly mutated**, another 237 mutated in two or more samples: grouped into 9 functional groups based on biologic function
- **7-gene-based subgroups in CN-AML** for disease prognosis by Marcucci et al., J Clin Oncol 2014
- **24-gene subgroups** for AML disease prognosis by Li at al., J Clin Oncol 2013
- leukemia stem cells and hematopoietic stem cells gene signatures for disease prognosis in CN-AML by Eppert et al., Nat Med 2011

Li-24-gene, Eppert-HSCR, Metzeler-86-probe and Bullinger-133-gene: significant overlap in gene lists (Wang et al., Leukemia 2017)

Gene expression profiling for prognosis in AML

- (Golub et al., Science 1999) and (Alizadeh et al. CSH Symp Quant Biol 1999) were the first who showed that **global gene expression profiling (GEP)** could be a tool for **cancer research and classification**
- Golub et al. proved that the distinction between two acute leukemias, AML, and ALL could be performed in a single test. Out of 6,817 human genes measured, **expression of 50 genes was selected as the most closely correlated with AML-ALL class distinction**

Each subtype of AML, including AML with chromosomal aberrations, mutations and cytogenetically normal AML, **possesses its own gene expression signature** and could be distinguished from one another (see Handschuh, J Oncology 2019 for a review)

Gerstung et al., Nat Comm 2015 showed that the **transcriptomic information provided the highest prognostic power among genomic, transcriptomic and clinical variables** (survival prediction, TCGA-AML cohort)

Prognostic panels (GEP) for risk stratification in AML

- **50 genes to distinguish between AML and ALL** by Golub et al., Science 1999
- **133-gene clinical-outcome predictor** based on gene expression profiles (N=116, including CN-AML) by Bullinger et al., N Engl J Med 2004
- 3 GEP-based clusters for AML prognosis by Valk et al., New Eng J Med 2004
- **21-gene predictor for subgroup classification** by Gutierrez et al., Leukemia 2005
- **GEP of 100 genes (SVM) → 6 groups of AML** by Haferlach et al., Blood 2005
- biggest GEP leukemia study: leukemia type classification (1400 probe sets with microarray technology) by Haferlach et al., Journ Clin Oncology 2010
- **40-marker gene classifer** (microarray probes) **for AML cytogenetic subtypes** by de la Bletiere et al., BMC Medical Genomics, 2012

Selection of the most “informative” genes

While gene expression profiling is clearly very informative for AML subtype differentiation, usually the gene selection or clustering algorithms are constructed with the outcome in mind (AML subtypes, cytogenetic abnormalities, treatment response, survival rates) and the number of clusters is chosen according to this context.

So, usually the lists of genes on the prev slides are the answer to the question “What’s the minimal set of genes to cluster these N samples to M groups of interest?”, while for a comprehensive splicing switch prediction the set of genes might be different. But as far as we’re ultimately trying to predict splicing changes in cancer-correlated genes (treatment-correlated?), then likely the prognostic GEP lists might be of use. How to curate this set otherwise?

Maybe we only take into account isoform switch genes, i.e. find out and try to predict AS-events effect on the splicing of the gene of interest? Along the way we check how it intersects with the GEP-predictive sets