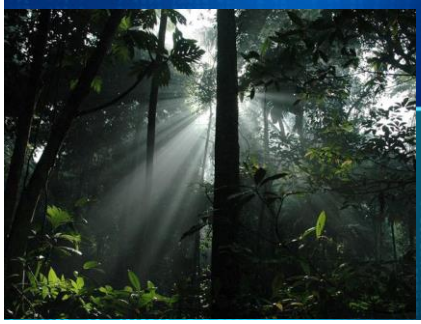
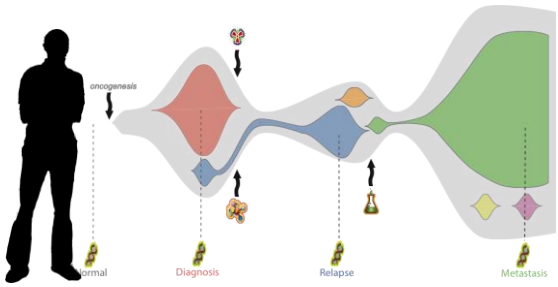
Forestial Problems and the Geometrization of Tumor Evolution



Raúl Rabadán



Comparing evolutionary histories

- Recently mutational trajectories of tumors have been studied in different context. Clinical/Biological questions:
  - Evolutionary questions:
    - how the different phases of tumors are related?
      - Type of evolutionary model:
      - Does it follow clonal replacement of dominant clone? Or only a few cells are responsible (niche, stem cells, etc)?
      - Is the dominant evolutionary factor high mutation rates or clonal heterogeneity?
      - how clones seeding metastasis are related?
  - Clinical questions:
    - » how informative (predictive value) is genomic information of primary tumor on later stages, therapy? Is it linear? Is it branched?
    - » how patients can be stratified using evolutionary genomic profiles?
  - Therapeutic questions:
    - » what are the specific molecular mechanisms driving relapse following certain therapies?

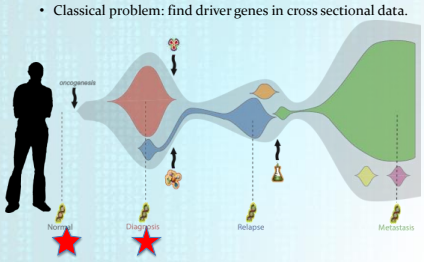
Cartography and landscapes

- Last 5 years have led to development of high throughput genomic technologies, leading to explosion of genomic data and methods to characterize the genomic profile of each tumor.
  - Currently more than 10,000 tumors have been profiled. Every tumor has different mutational spectrum, result of different evolutionary history.
- If every tumor follows a different evolutionary path, how do we go beyond cartography and case reports?
  - Main path: Cross sectional studies.
  - Problem: define the "metric space" where the stories can be compared.
  - Two key ingredients: **metrics** and **statistics**.
  - Most strategies based on convergent evolution, i.e. recurrence: find common alterations, pathways, etc that are altered in statistically significant sets of patients.

HOW CAN WE COMPARE EVOLUTIONARY HISTORIES?

Comparing evolutionary histories

- Previous studies (including large studies TCGA, and our own) characterize single sample per patient. Methods to extract information from cross-sectional studies have been developed and applied in this context.
- Classical problem: find driver genes in cross sectional data.

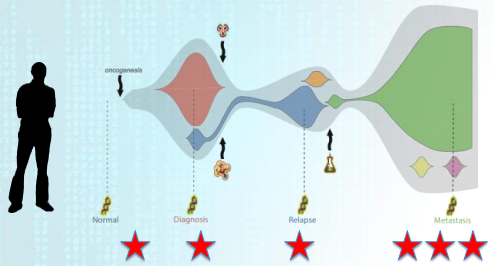


Frattini et al, *Nature Genetics*
Singh, et al *Science*,
Tiacci et al , NEJM
Pasqualucci et al, *Nature*,
Palomero, *Nature Genetics*,
Wang, *eLife*,
Tzoneva, *Nature Medicine*,

Etc.

Comparing evolutionary histories

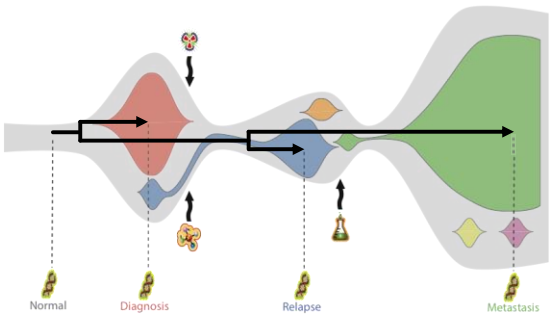
- Previous studies (including large studies TCGA) characterize single sample per patient. Methods to extract information from cross-sectional studies have been developed and applied in this context.



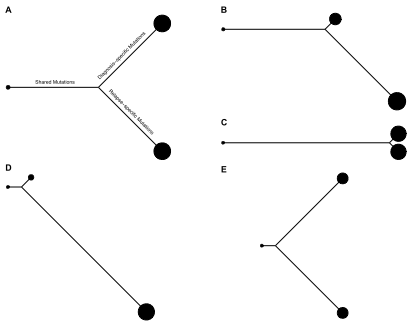
Comparing evolutionary histories

- If every tumor follows a different evolutionary path, how do we extend computational and statistical frameworks to study longitudinal data?
- Aims:
  1) to develop a framework to compare the evolutionary history of different patients -> **metric space** that allow comparison of evolutionary histories.
  2) To develop a **statistical framework** and **machine learning approaches** to longitudinal cohorts.
  3) To develop a statistical/evolutionary framework **to test different evolutionary models** (e.g. multistage due to hard selective sweeps).

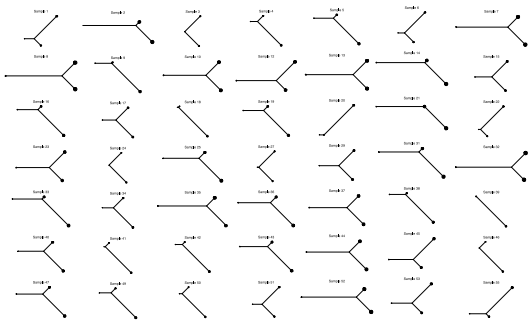
Cancer follows clonal Darwinian evolution



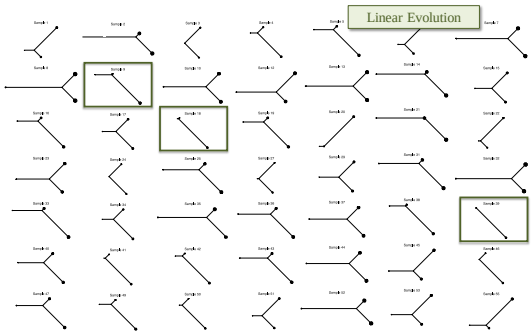
Classifying Evolutionary Behaviors



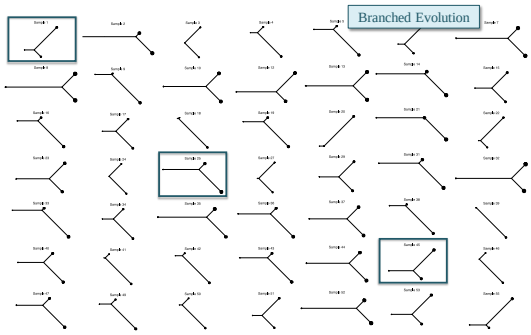
Patients represent forests of phylogenetic trees



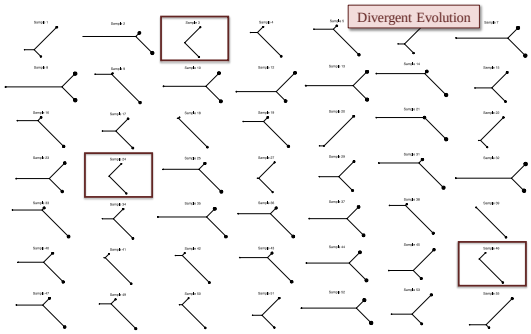
Patients represent forests of phylogenetic trees



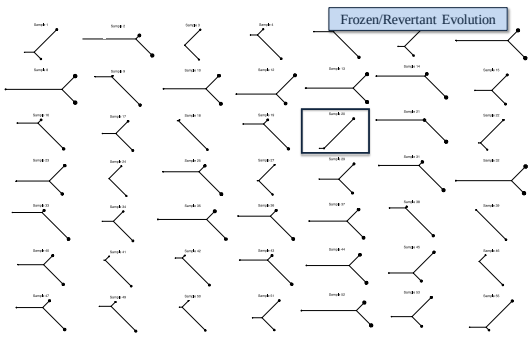
Patients represent forests of phylogenetic trees



Patients represent forests of phylogenetic trees



Patients represent forests of phylogenetic trees



Forestial Questions

Interesting properties I would like to capture:

1) I want to compare trees: how similar are two trees? **Distance**.

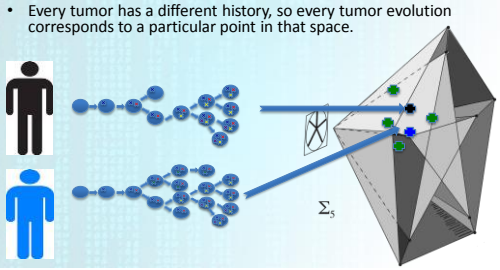
2) I want to do **statistics**: means (average tree), centroids and variances.

- I want to apply **machine learning** techniques to uncover evolutionary patterns.

3) I want to **test different evolutionary models** (e.g. multistage hard selective sweeps).

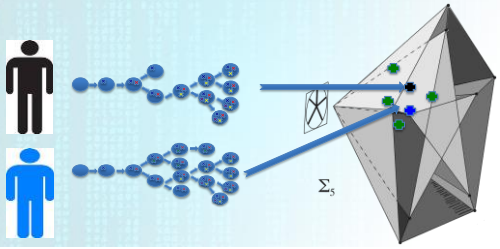
Moduli space of clonal evolution

- To compare two or more histories we would like to find a space that describe all possible evolutionary histories.
  - Moduli space: space whose points correspond to objects of some kind.
- Every tumor has a different history, so every tumor evolution corresponds to a particular point in that space.



Moduli space of clonal evolution

- Instead of talking of:
  - A tumor history -> point in a space
  - All potential histories -> moduli space
  - Cross-sectional data -> distributions of points in moduli space.



Space of trees

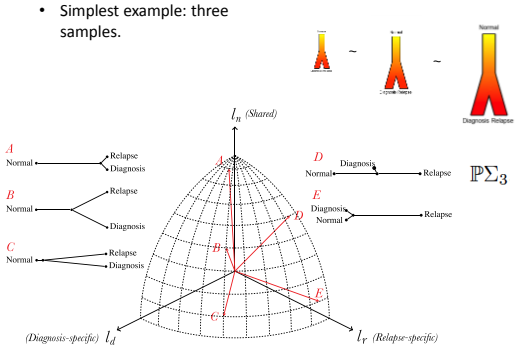
Theorem 1: The space of trees is a Polish space (one can define statistics).

Theorem 2: The space of trees is a CAT(0) space (one can define means and variances in a meaningful way).

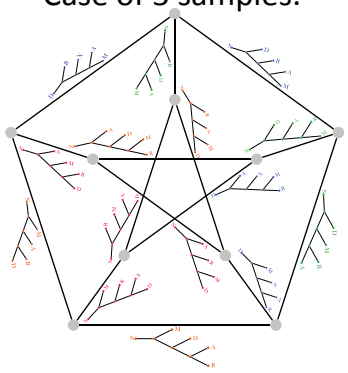
Billera, Holmes, Vogtman (2003)
Zairis, Khiabanian, Blumberg, Rabadan (2014)

Moduli projective space of clonal evolution

- Simplest example: three samples.



Case of 5 samples.

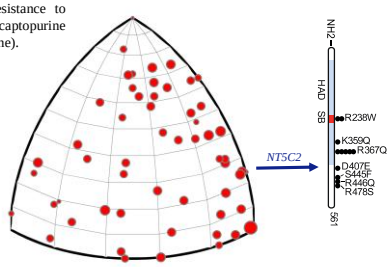


Relapse-specific mutations in ALL

- *NT5C2* mutations in ALL cells induce resistance to therapy (6-mercaptopurine and 6-thioguanine).

Collaboration with T. Palomero and A. Ferrando.

H. Khiabanian



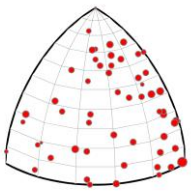
Tzoneva *et al*, Nat. Medicine 2013

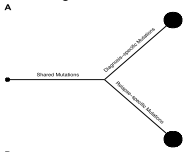
CAN WE RECONSTRUCT THE MAIN ROUTES OF TUMOR EVOLUTION?

Pediatric leukemias (ALL) at relapse: highly branched process

Highly branched process:
- dominant clone at diagnose is erradicated.
- ancestral clone adquires additional mutations that confer resistance to therapy.

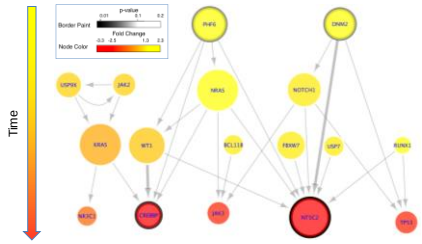
Branched evolution allows to trace back some mutations:
- Reconstruct ancestral clone,
- order the chronological order of mutations.





J. Wang et al, *eLife*

Pediatric leukemias (ALL) at relapse: highly branched process



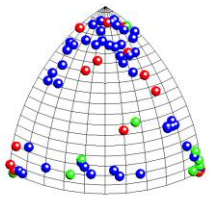
Early events:
- PHF6 (*Nat. Genetics).*

Late events associated to relapse:
- NT5C2 driven relapses (*Nat. Medicine*).

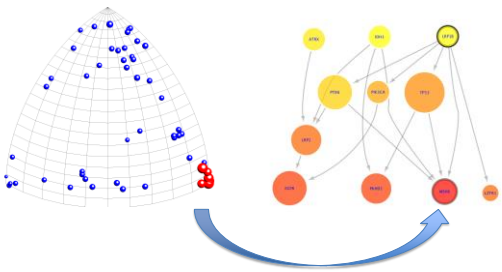
GBM evolution shows distinct evolutionary trajectories associated to specific molecular mechanisms

Collaboration with A. Iavarone and G. Finocchiaro.
J. Wang.

Three data sets:
- ours,
- TCGA triplets,
- Verhaak et al.

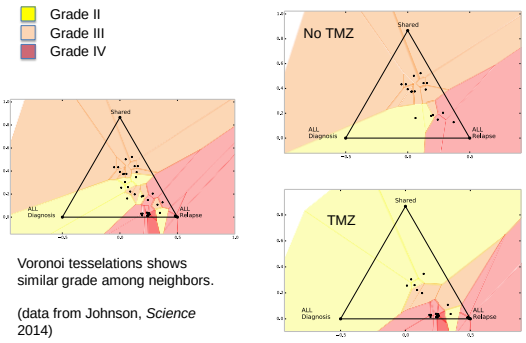


GBM evolution shows distinct evolutionary trajectories associated to specific molecular mechanisms



DOES EVOLUTIONARY HISTORY CORRELATES WITH PROGNOSIS?

Clinical Correlates: Grade of Low Grade Glioma at recurrence 's is associated with position in moduli space

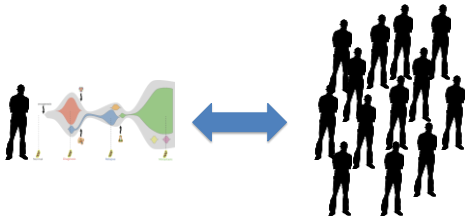


Voronoi tesselations shows similar grade among neighbors.

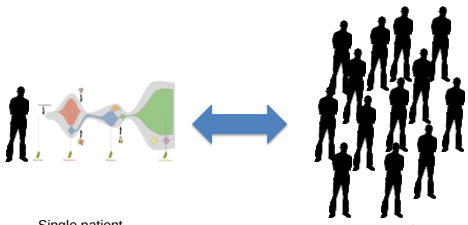
(data from Johnson, *Science* 2014)

ARE THERE PATTERNS IN PROGRESSION OF TUMORS?

Longitudinal vs Cross-sectional



Longitudinal vs Cross-sectional



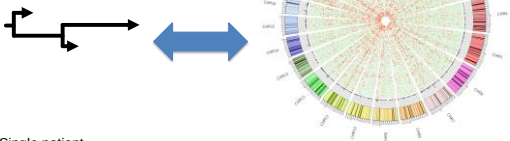
- Single patient
- Evolution (time data)

- Evolutionary language:
  - Fitness
  - phylogenies

- Many patients
- No evolution

- Statistical language:
  - Frequency of alterations

Longitudinal vs Cross-sectional



- Single patient
- Evolution (time data)

- Evolutionary language:
  - Fitness
  - phylogenies

- Many patients
- No evolution

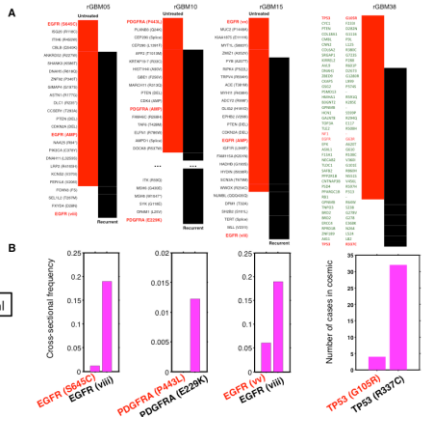
- Statistical language:
  - Frequency of alterations

Principle of common clonal replacement

If in the evolution of a tumor a clone is replaced with another one with different alterations in a particular gene, then the second alteration is more common than the first.

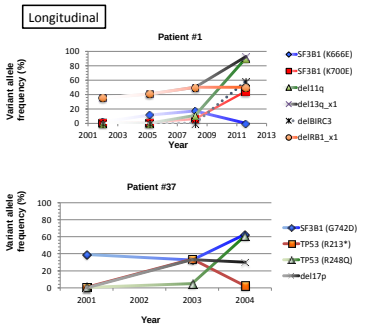
Longitudinal

Cross-sectional

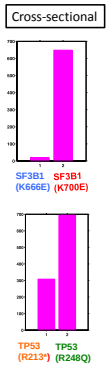


CLL

Longitudinal

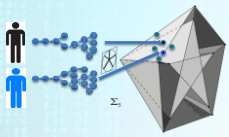


Cross-sectional



Questions

- Can we compare evolutionary cancer histories?
  - CAT(0) spaces: allow define well-behave statistics.
- Can we reconstruct main evolutionary routes?
  - Translate longitudinal data into directed graphs.
- Does evolutionary histories correlate to clinical information?
  - CAT(0) spaces: allow machine learning approaches.
- Are there patterns of progression?
  - Clonal replacement by "cross-sectional" frequent mutations.



Columbia University

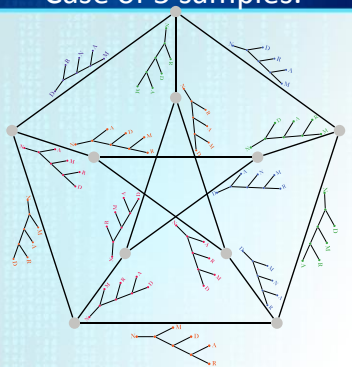
Center for Topology of Cancer Evolution and Heterogeneity

A member of the National Cancer Institute's Physical Sciences in Oncology Network

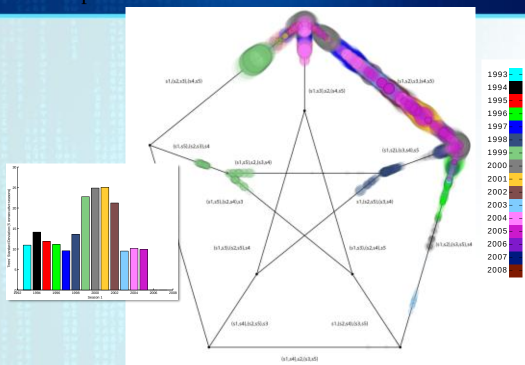
- National Cancer Institute (NCI).
- Leukemia and Lymphoma Society.
- Lymphoma Research Foundation.
- Geneva Foundation.
- Stewart Foundation.

- Columbia University:
  - Departments of Systems Biology and Biomedical Informatics:
    Hossein Khiabanian, Jiguang Wang, Oliver Elliott, Pablo Camara, Rachel Melamed, Albert Lee, Francesco Abate, Sakellarios Zairis, Kevin Emmett, Daniel Rosenbloom, Chioma Madubata.
  - Institute for Cancer Genetics and Irving Cancer Center:
    R. Dalla Favera, A. Ferrando, T. Palomero, L. Pasqualucci, G. Fabbri, A. Iavarone, A. Lasorella, J. Celebi, S. Mukherjee.
- Texas (Austin), MSRI (Berkeley): **A. Blumberg.**
- Stanford: G. Carlsson.
- Milan: **G. Finnochiaro**

Case of 5 samples.

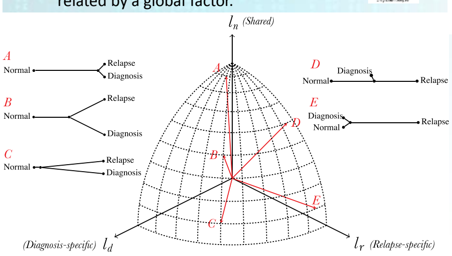


Space of trees with 5 consecutive seasons



Moduli projective space of clonal evolution

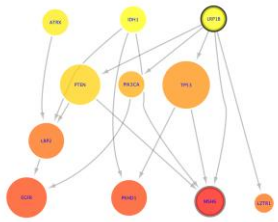
- A way of representing and compare the evolution of tumors in different patients is to identify histories that are related by a global factor.



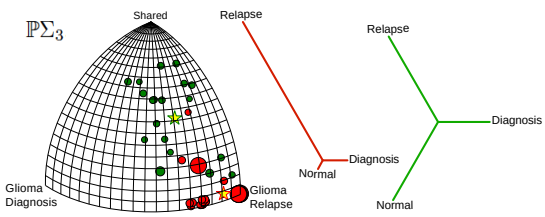
Funding:
- National Cancer Institute (NCI).
- National Institute on Alcohol Abuse and Alcoholism (NIAAA).
- National Library of Medicine (NLM).
- Leukemia and Lymphoma Society.
- Lymphoma Research Foundation.
- Geneva Foundation.
- Stewart Foundation.

- Columbia University:
  - Depatments of Systems Biology and Biomedical Informatics:
    Hossein Khiabanian, Jiguang Wang, Oliver Elliott, Pablo Camara, Rachel Melamed, Albert Lee, Francesco Abate, Sakellarios Zairis, Kevin Emmett, Daniel Rosenbloom, Chioma Madubata.
  - Institute for Cancer Genetics and Irving Cancer Center:
    R. Dalla Favera, A. Ferrando, T. Palomero, L. Pasqualucci, G. Fabbri, A. Iavarone, A. Lasorella, J. Celebi, S. Mukherjee.
- Texas (Austin), MSRI (Berkeley): **A. Blumberg.**
- Stanford: G. Carlsson.
- Roma: **G. Finnochiaro**

GBM evolution shows distinct evolutionary trajectories associated to specific molecular mechanisms

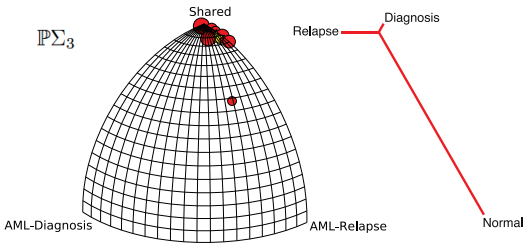


Glioma's therapy-driven highly branched evolution: TMZ drives hypermutation



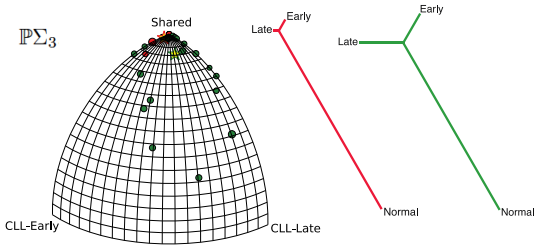
Evolution to low grade to high grade glioma (Johnson, *Science* 2014)

Acute myelogenous leukemia: frozen evolution



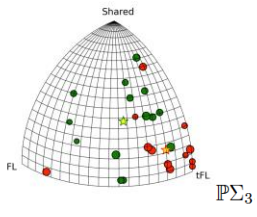
Relapsed AML (Ding et al, Nature 2012): 8 AML patients. Standard induction chemotherapy for AML (anthracycline and cytarabine). Chemotherapy failed to eradicate the founding clone.

Linear vs branched: Chronic lymphocytic leukemia untreated vs. treated: clonal replacement and classic hard selective sweeps



(Landau, *Cell* 2014)

Branched transformation of follicular lymphoma to DLBCL



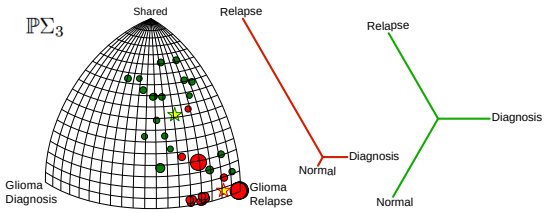
Follicular lymphoma (FL) is an indolent disease, but 30%-40% of cases undergo histologic transformation to an aggressive malignancy

- Red dots: Collaboration with R. dalla Favera and L. Pasqualucci (Cell Reports 2014).
- Okusun (Nat Genetics 2014). 10 triplets.
- Summary:

  - branched transformation: FL is not a direct ancestor of tDLBCL.

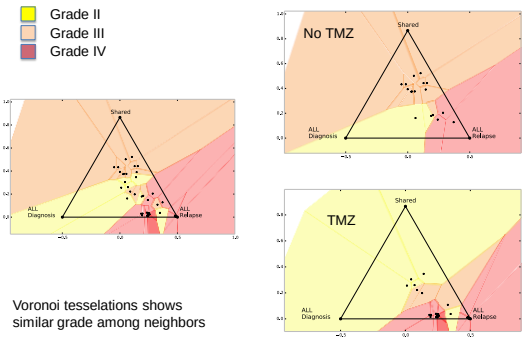
  - long branch leading to tFL associated with alterations deregulating cell-cycle progression and DNA damage responses (CDKN2A/B, MYC, and TP53)

Glioma's therapy-driven highly branched evolution: TMZ drives hypermutation



Evolution to low grade to high grade glioma (Johnson, *Science* 2014)

Grade of Low Grade Glioma at recurrence 's is associated with position in moduli space

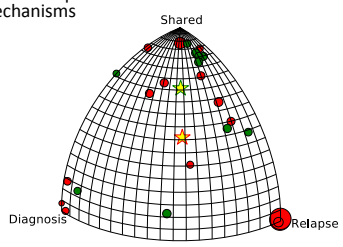


Voronoi tesselations shows similar grade among neighbors

GBM evolution shows distinct evolutionary trajectories associated to specific molecular mechanisms

Collaboration with A. Iavarone and G. Finocchiaro.

J. Wang.

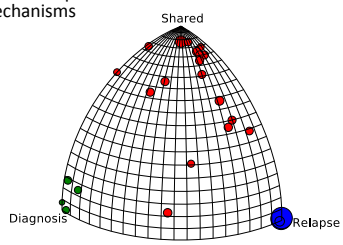


GBM evolution shows distinct evolutionary trajectories associated to specific molecular mechanisms

Collaboration with A. Iavarone and G. Finocchiaro.

J. Wang.

Three clusters.

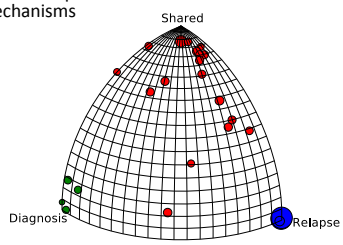


GBM evolution shows distinct evolutionary trajectories associated to specific molecular mechanisms

Collaboration with A. Iavarone and G. Finocchiaro.

J. Wang.

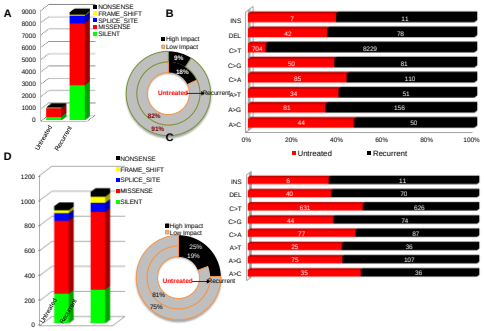
Three clusters.



HMC:

- MSH6 mutations (p-value 0.01)
- Transcription associated hypermutation, C->T.

GBM evolution: hypermutation cluster

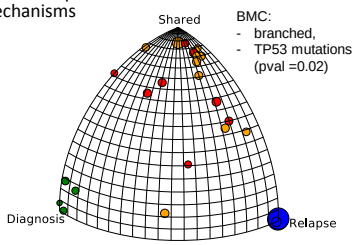


GBM evolution shows distinct evolutionary trajectories associated to specific molecular mechanisms

Collaboration with A. Iavarone and G. Finocchiaro.

J. Wang.

Three clusters.



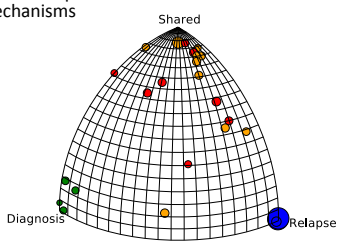
BMC:
- branched,
- TP53 mutations (pval =0.02)

GBM evolution shows distinct evolutionary trajectories associated to specific molecular mechanisms

Collaboration with A. Iavarone and G. Finocchiaro.

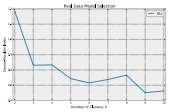
J. Wang.

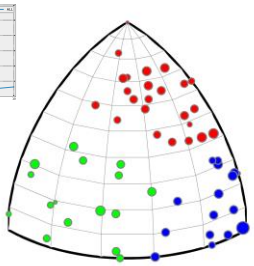
Three clusters.



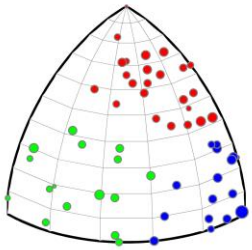
LMC:
- Low mutation cluster at relapse, ???

Pediatric leukemias (ALL): 3 evolutionary trajectories





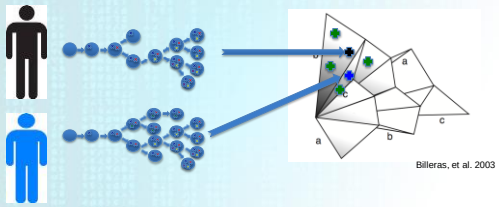
Pediatric leukemias (ALL)



NT5C2 driven relapses
Nat. Medicine

Moduli projective space of clonal evolution

- The extension of the previous examples for cases with more longitudinal data.
- The moduli space with m branches is extremely interesting.
- They have to do with the space of unrooted phylogenetic trees.



Billeras, et al. 2003

Moduli projective space of clonal evolution

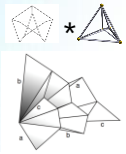
- Moduli space is the Cartesian product of two spaces, providing information on:
  - **Topology**: l_m = Internal branches described by Billeras, Holmes and Voigtmann = BHV_{m-1}
  - **Linearity**: E_m = External branches by positive quadrant in R^m.
  - Important property: **CAT(0)**.
- Projective Moduli space is the join of two spaces (allows to compare different experiments):
  - Internal branches: Boardman space,
  - External branches: Simplex in m-1 dimensions.
  - Important property: exponential branch length distribution translates into uniform distribution in projective coordinates.

$$\Sigma_m = BHV_{m-1} \times \mathbb{R}^{m+}$$

$S_m = l_m \times E_m$
$\dim(l_m) = m - 3$
$\dim(E_m) = m$

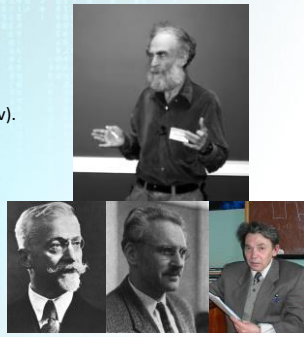
$$\mathbb{P}\Sigma_m = \tau_{m-1} \star T_{m-1}$$

t_{m-4} = Boardman space of dimension m - 4,
T_{m-1} = simplex of dimension m-1.

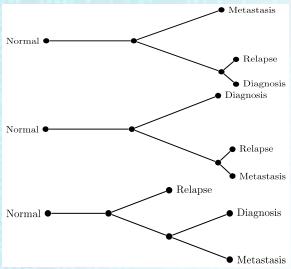


CAT(0) spaces

- *Theorem 1*: The space of trees is CAT(0).
- Coined by Gromov in 1987 (acronym for Cartan, Aleksandrov and Toponogov).
- They are spaces where geodesics can be uniquely defined, one can define centroids, variances, statistics, etc.

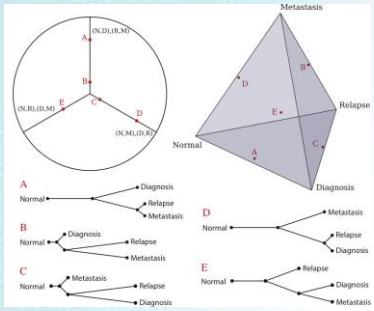


Forestial problem: tree topologies



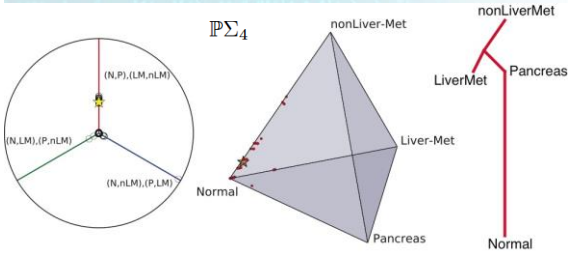
Moduli projective space of clonal evolution

- Four samples: 4 leaves could have 3 topologies.



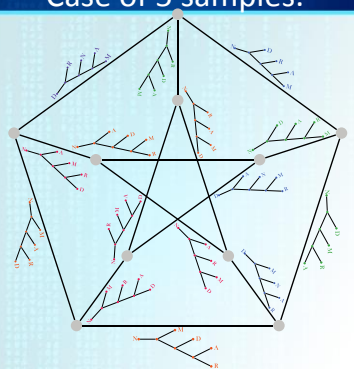
$\mathbb{P}\Sigma_4$

Metastatic Pancreatic Cancer



- Metastasis of pancreatic tumors to the liver and beyond. Question: what is the source of metastasis? (Campbell et al Nature 2010).
- Metastasis derive linearly from primary.

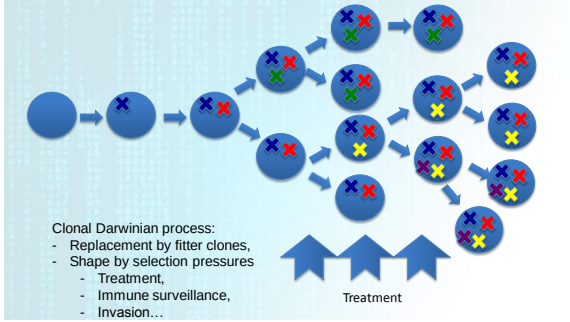
Case of 5 samples.



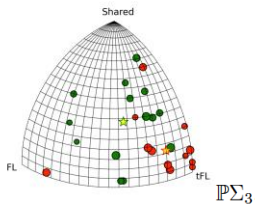
More complex evolutionary behaviors

Underlying evolutionary mechanism	Large N eff	Small Neff: few stems cells, niche, residual cells, etc
No mutation: epigenetic effects?	Frozen evolution	Frozen evolution
Hard selective sweeps: a *de novo* mutation with strong selective advantage.	Linear evolution: **mutation rates** are most dominant effect in relapse.	Branched evolution: **clonal heterogeneity** in small population.
Soft selective sweeps: pre-existing alleles (standing genetic variation) are selected under novel selective pressures.	Branched evolution: **clonal heterogeneity** in large population.	Branched evolution: **clonal heterogeneity** in small population.
Negative selection: elimination of pre-existing clones with alterations that make sensitive to therapy.	Linear evolution: aggressiveness and duration of therapy.	Branched evolution.
Iatrogenic induced relapse.	Divergent evolution.	Divergent evolution.
Polygenic adaptation.	Linear evolution: **mutation rates** are most dominant effect in relapse.	Branched evolution: **clonal heterogeneity** in small population.

Standard model of Cancer Evolution



Branched transformation of follicular lymphoma to DLBCL



Follicular lymphoma (FL) is an indolent disease, but 30%-40% of cases undergo histologic transformation to an aggressive malignancy

- Red dots: Collaboration with R. dalla Favera and L. Pasqualucci (Cell Reports 2014).
- Okusun (Nat Genetics 2014). 10 triplets.
- Summary:

- branched transformation: FL is not a direct ancestor of tDLBCL.

- long branch leading to tFL associated with alterations deregulating cell-cycle progression and DNA damage responses (CDKN2A/B, MYC, and TP53)