6.872 / HST650
A computational modeling view of cancer biology and development

Informatics Program / *HST-Children's Hospital Boston*

Pediatric Oncology / *Dana-Farber Cancer Institute*

Alvin T. Kho

Thursday, 07 October 2004

Outline

SYSTEM

- Biological background & motivating ideas, questions
- Our case study/problem:
  - Medulloblastoma – a human cerebellar tumor
  - Cerebellar development in the mouse
  - Biological question/s

MODEL

- Experiment design, measurement platforms & data
- Modeling data. Mathematical formulation of biological question/s

MODEL

- Results of model
  - Molecular & genomic correlates
  - Metastasis & development time windows

SYSTEM

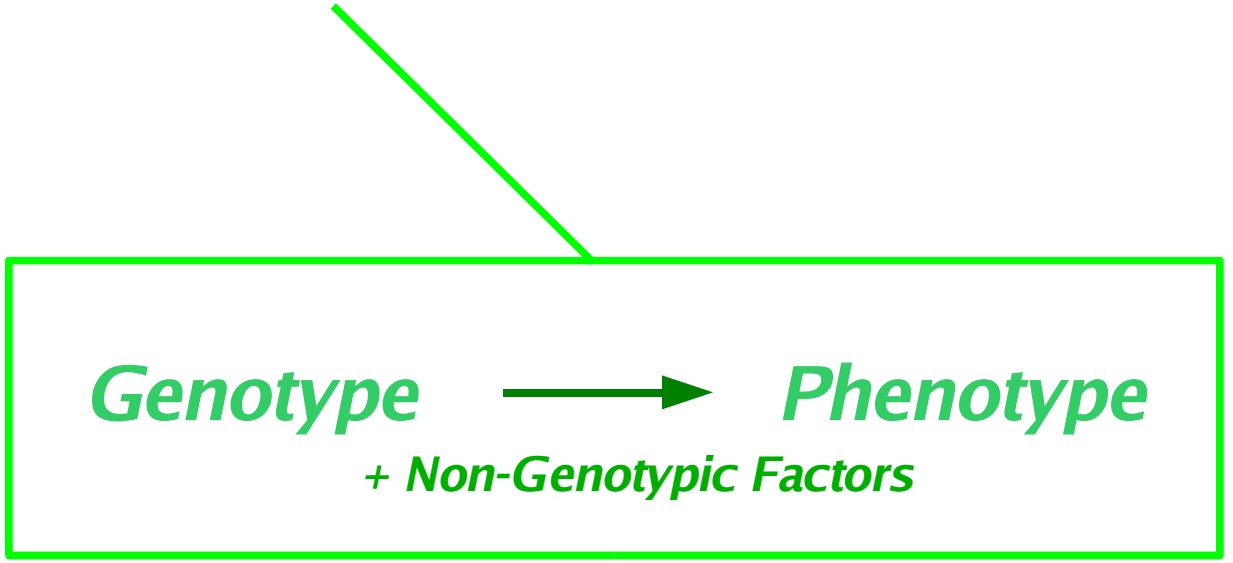
- Translating math model outcomes into testable hypotheses in biology & biological knowledge
- Reality checks, discussion, questions

Biological background & motivating ideas, questions

- ***What is cancer?*** General term for >100 diseases characterized by abnormal / uncontrolled growth of cells. Cancer cells can invade, destroy surrounding normal tissues. From its original site, it can also spread via blood / lymph system to establish new cancers in distant tissue. Tumor = mass of cancer cells.
  - **Basic cancer biology question:** What are the (molecular) mechanisms that underwrite the formation and natural history of a cancer?
  - **Basic medical question:** How do we destroy cancer – minimizing negative effects on the patient?
- ***What is development?*** The process of cellular differentiation that leads to the formation of a mature biological system, e.g., an organ (morpho-/organo-genesis), an organism (embryogenesis).
  - **Basic question:** What are the (molecular) mechanisms that underwrite the developmental process?

Biological background & motivating ideas, questions

- ***Biological Meta Question*** Unravel the mechanisms and physical-temporal scales involved in this transformation



SCALES . . .

Microscopic		***Macroscopic***
Gene		Organism
RNAi	***"Environmental"***	Tissue
Protein		Behaviour
Phosphorylation cascade		Disease symptom
Metabolic network		Treatment response

Biological background & motivating ideas, questions

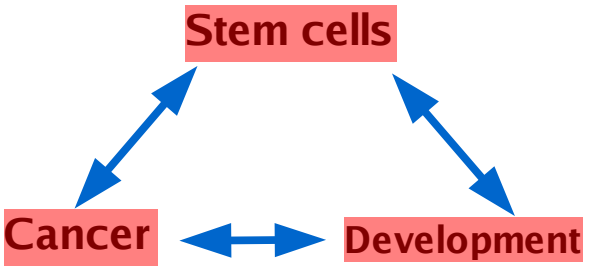
- ***An old observation:*** Tumors and corresponding early developing tissue / cells share certain common features such as
  - Simple (unorganized) morphology
  - Potential to differentiate into various cell types
  - Capacity for extensive growth

Univ. Manitoba Health Sciences Lib

- ***Embryonal rest theory*** (J. Cohnheim 1875): Cancer = abnormal activation of embryonic "vestiges" that generally lie dormant in (adult) tissues.
  - Cell-of-Origin questions. Where / How does cancer arise?
  - Is there common biology / mechanisms underlying tumorigenesis and development?
  - Which early developing tissue? What is a relevant development framework? How do we determine it? Is this framework unique?

Biological background & motivating ideas, questions

- ***What have we learned since the 19th century?***
  - ***Cancer informs Development*** Molecules discovered based on their role in cancer – oncogenes, tumor suppressors – are fundamental regulators of cell growth & differentiation during development
  - ***Development informs Cancer*** Molecules identified as regulators of morphogenesis have been implicated in cancer biology
- Developmental basis for "liquid" cancers well known – e.g., lymphomas, leukemias. Solid tumors? Not clear.
- CNS tumor classification today is based on its histologic resemblance to cells of the developing CNS (Bailey & Cushing, 1925).
- 21st century version of *embryonal rest theory:*
  Cancers derive from "stem"/progenitor cells



Biological background & motivating ideas, questions

- ***Our question/s***

  1. How "similar" is cancer to different developmental stages on a genomic (over many many molecules) scale?

  2. ***How shall we define "similarity" between two biological systems?***

  3. How will establishing / disproving this "similarity" advance the current state-of-knowledge in cancer biology?

- ***Our case study thread for this lecture ...***

  - Human medulloblastoma

  - Mouse cerebellar development

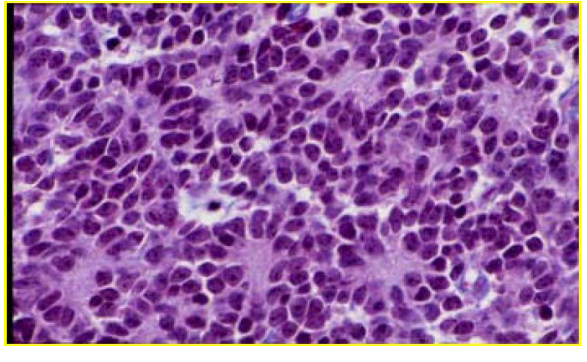
ClipArt.com

Losing forest for the trees ...

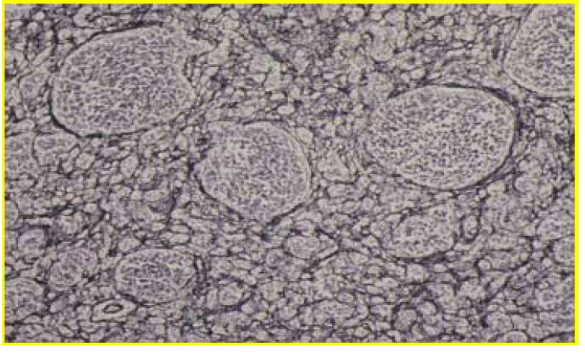
Medulloblastomas (MB) in human

- ***MB.*** A primitive neuroectodermal tumor of the cerebellum with predominant neuronal differentiation. Two major variants:
  - ***Classic.*** Undifferentiated "small blue" malignant cells with round/oval hyperchromatic nuclei
  - ***Desmoplastic.*** ~20-30% of MBs. Nodules of localized neuronal differentiation in cellular sea of undifferentiated malignant cells. Abundant reticulin, collagen network (dark)
- Most common pediatric malignant brain tumor 30%. 2:100,000 kids ~350 new cases in U.S. per year. Median age 9. M:F = 3:2, poorer prognosis for M.
- A feature of genetic disorders: Gorlin (Chr:9q PTCH), Von Hippel-Landau (ChrL3p, 11q, VHL), Turcot (Chr:5q, APC).
- Genetic factors affect MB formation. Most common, Chr:17p loss.
- Rx: Surgery + radio/chemotherapy. Recurrence & metastasis (to CNS and bone) common. 5-year patient survival ~50%

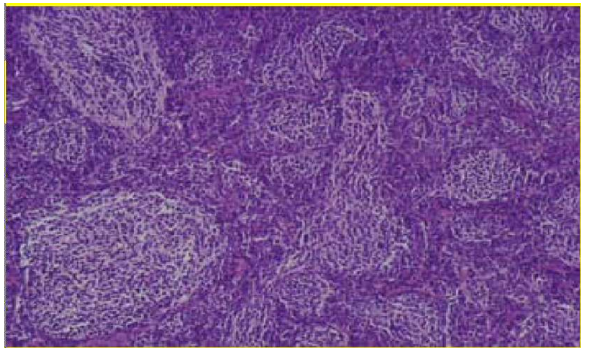
Atlas Interactif de Neuro-Oncologie, Univ. Nice Sophia-Antipolis



c MB

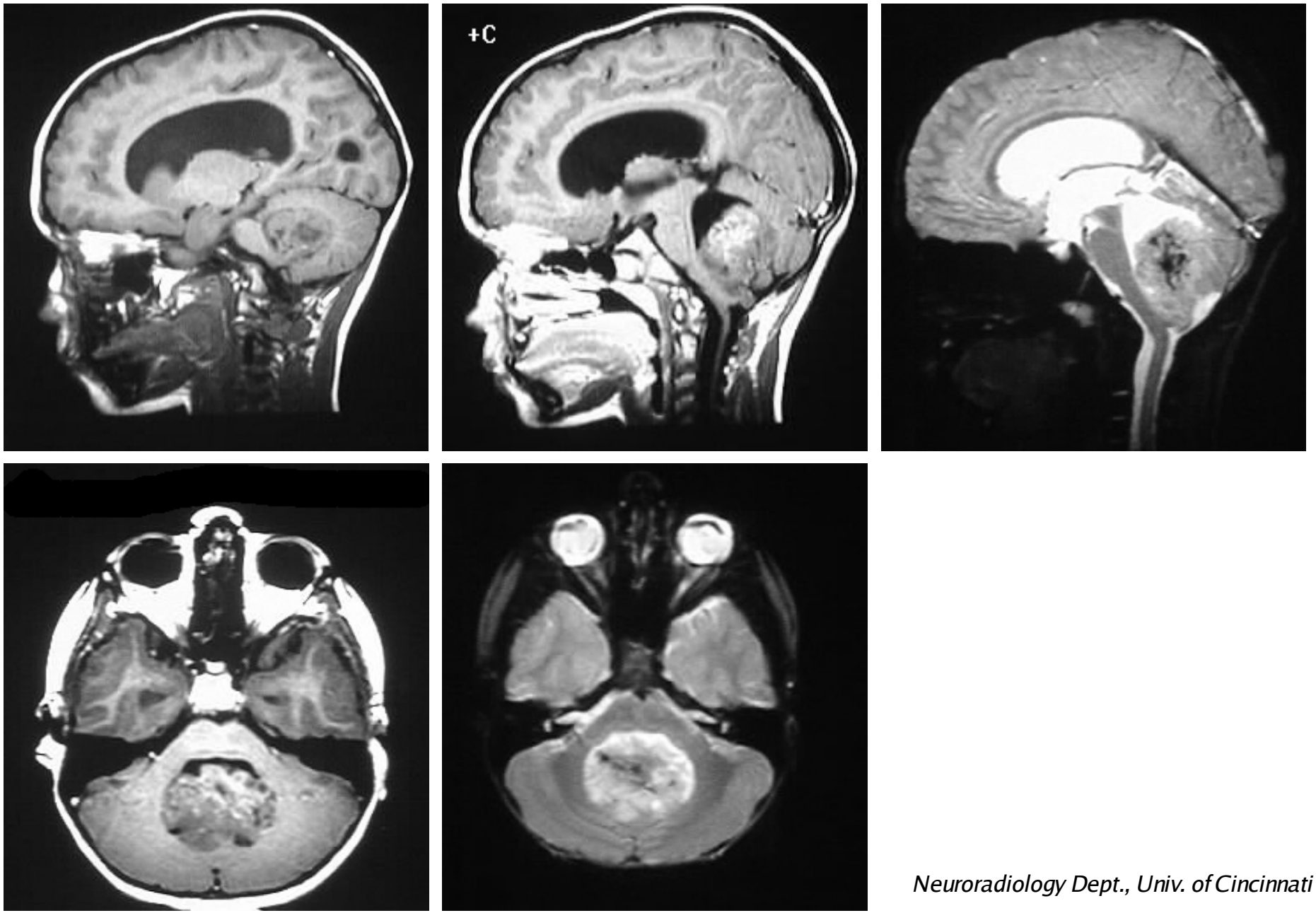


d MB



d MB

Medulloblastomas (MB) in human



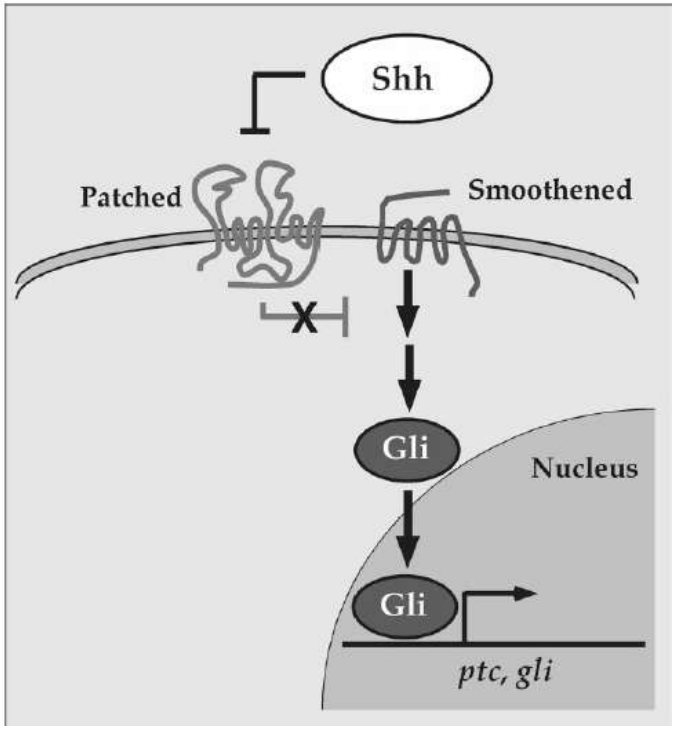
Neuroradiology Dept., Univ. of Cincinnati

Medulloblastomas (MB) in human

- ***CGNP origin?*** MB cells not consistent with normal cerebellar cells, but most similar to cerebellar granule cell precursors/CGNP. Expresses CGNP-specific transcription factors: *ZIC*, *NSCL1*. Low occurrence of *P53* mutation

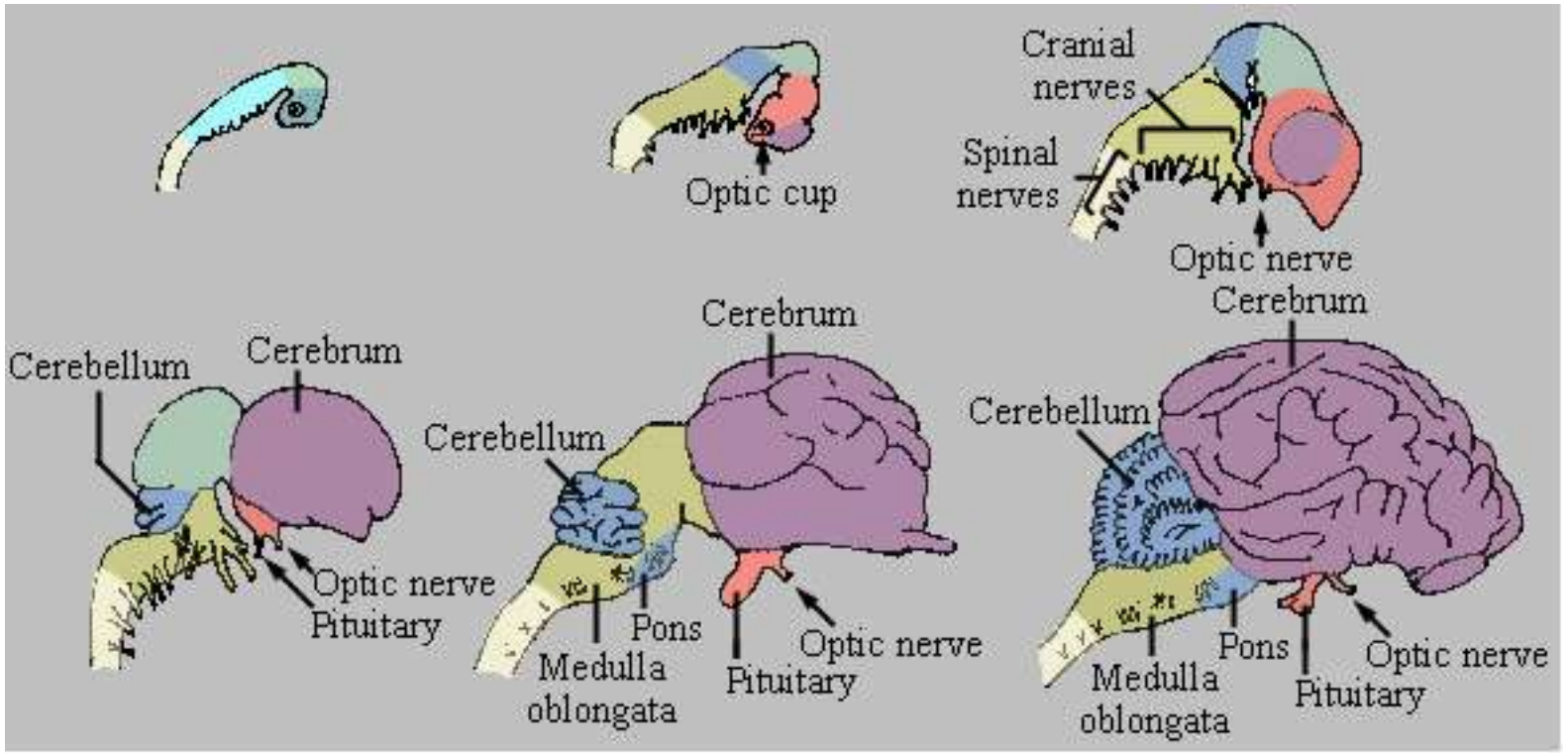
- Sonic Hedgehog/SHH morphogen – fruitfly embryo segmentation, CNS patterning.

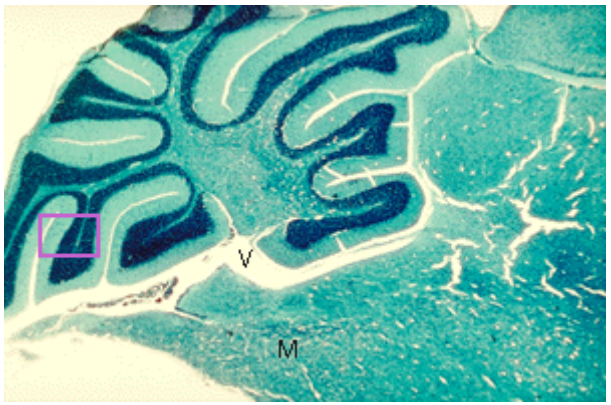
- *Patched/Ptch*+/- mice. 15-25% develop cerebellar tumors - small round cells on cerebellar surface morphologically identical to human MB. (Also rhabdomyosarcomas)

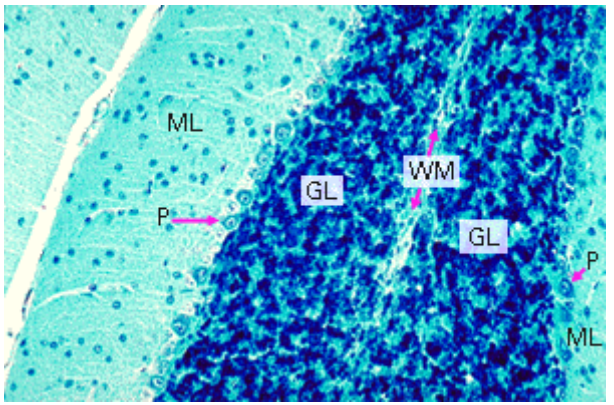


Wechsler-Reya & Scott, Annu. Rev. Neurosci 2001

Cerebellar (Cereb) development in mouse



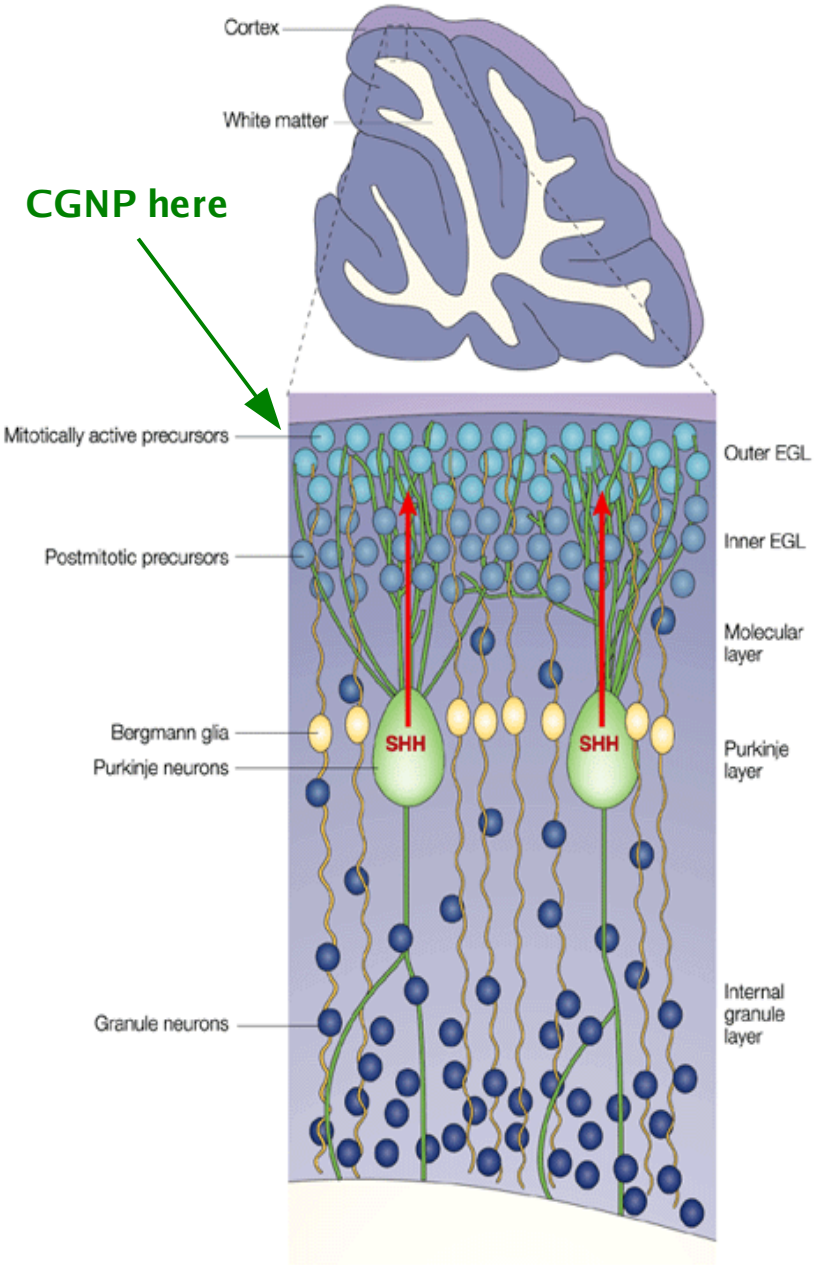




Cell & Developmental Biology Online, Univ. of Guelph

Cerebellar (Cereb) development in mouse

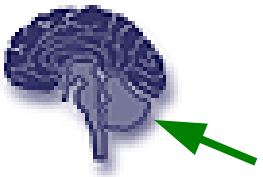
- ***Function.*** Motor coordination, physical skill memory
- ***Postnatal development.*** Granule cells/CGNP proliferation on external layer P1-7. Migration to internal layer. Differentiation >P7. Mature ~P30.
- ***Layer Architecture.*** Correlation between layers and development stage.
- ***Granule Layers.*** EGL to IGL.
- ***Molecular Markers.*** Layer/Stage-specific.
- ***Evolution.*** Conserved growth mechanism across species.



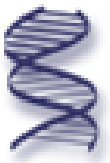
Altaba et al., 2002 Nature Reviews | Neuroscience

Experiment design, measurement platform & data

- ***Human samples***
  - 4 normal cerebellar tissue
  - 9 d MB
  - 22 c MB



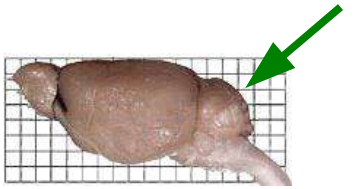
- ***Human "control" samples***
  - 17 normal lung tissue
  - 21 squamous cell lung carcinoma
- ***RNA measurement platform***



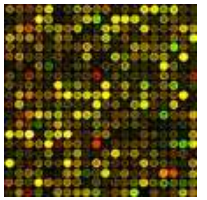
  - Microarray, Affymetrix Hu6800
  - 5,501 unique LocusLink ID'ed genes

- ***Mouse samples***

  - Postnatal whole cerebellum: P1, 3, 5, 7, 10, 15, 21, 30, 60
  - Duplicated



- ***Mouse "control" samples***
  - Whole lung: E12, 14, 16, 18; P1, 4, 7, 14, 21, AD
  - No replicates
- ***RNA measurement platform***

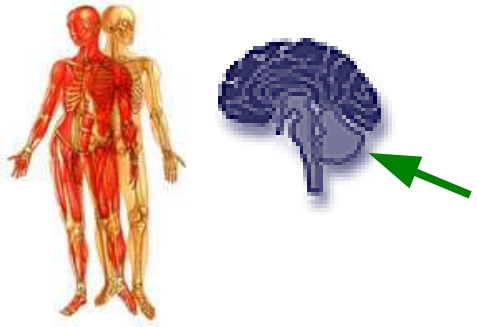


  - Microarray, Affymetrix Mu11K
  - 7,577 unique LocusLink ID'ed genes

Images: NIH, NCBI, JAX, MBL

Experiment design, measurement platform & data

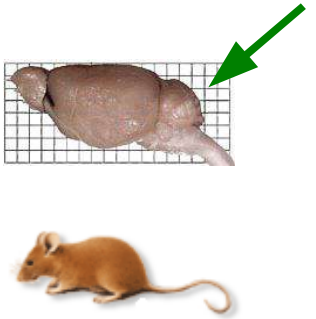
Gene/DNA → *RNA* → *Protein*



HUMAN GENE	SIGNAL
BACH1	63.1
ABCC6	14.7
CASP8	21.0
HOXA5	50.3
...	...

Cross-species gene matching

HomoloGene @NCBI



MOUSE GENE	SIGNAL
Bach1	126.1
Abcc6	29.4
Casp8	42.0
Hoxa5	100.6
...	...

Orthologous pairs of molecules are matched via calculation (inferred from sequence) or curated from shared functional features. *Not a 1-1 mapping!*

3,477 unique genes common to both species assay platform

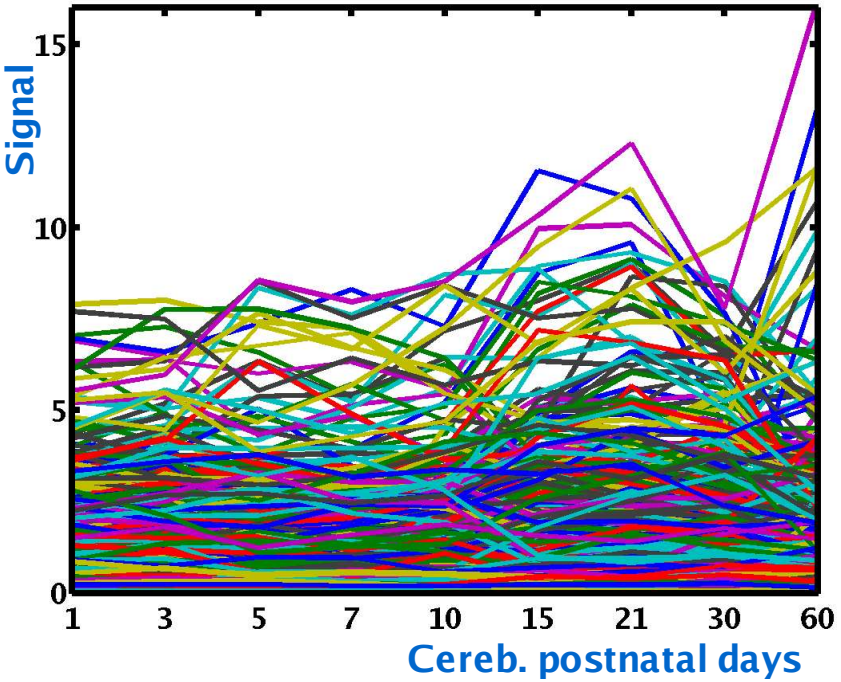
Images: NIH, NCBI, JAX, MBL

Modeling data, mathematical formulation/s of our problem

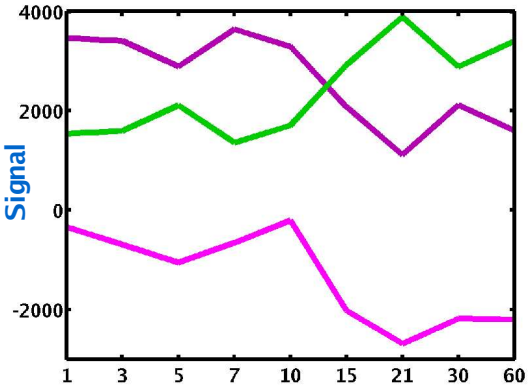
- ***Data representation/s,*** 2 views: (... recalling earlier multi-scalar property of biological phenomena)
  - ***Global/"Genomic":*** Each human / mouse sample is represented as a vector of N reported RNA levels. N, a positive integer = # of different genes / RNA species.
  - ***Molecular/"Gene-by-gene":*** Each gene is a vector of D reported expression measurements. D, a positive integer = # of conditions in which gene was assayed.
- ***Definition/s of "similarity,*** Assessing "similarity" between 2 mathematical objects, e.g., vectors:
  - ***Define a metric:*** Or more weakly, a measure of similarity. Does this metric definition agree / conflict with the biological notion of similarity in the current study? E.g., Pearson correlation, Euclidean distance

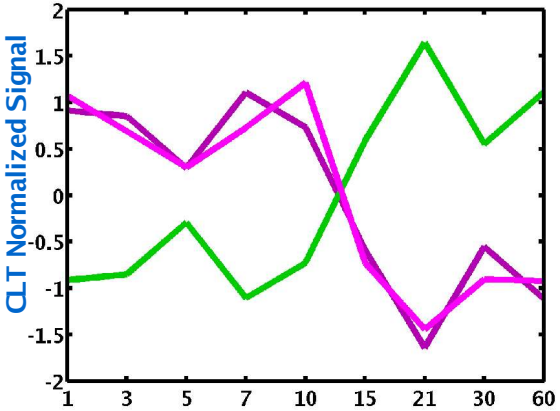
	Exp_1	Exp_2	Exp_3	. . .
Gene_1	*x11*	*x12*	*x13*	. . .
Gene_2	*x21*	*x22*	*x23*	. . .
Gene_3	*x31*	*x32*	*x33*	. . .
. . .	. . .	. . .	. . .	. . .

Modeling data: Visualizing time profiles of individual genes during mouse cereb. dev.

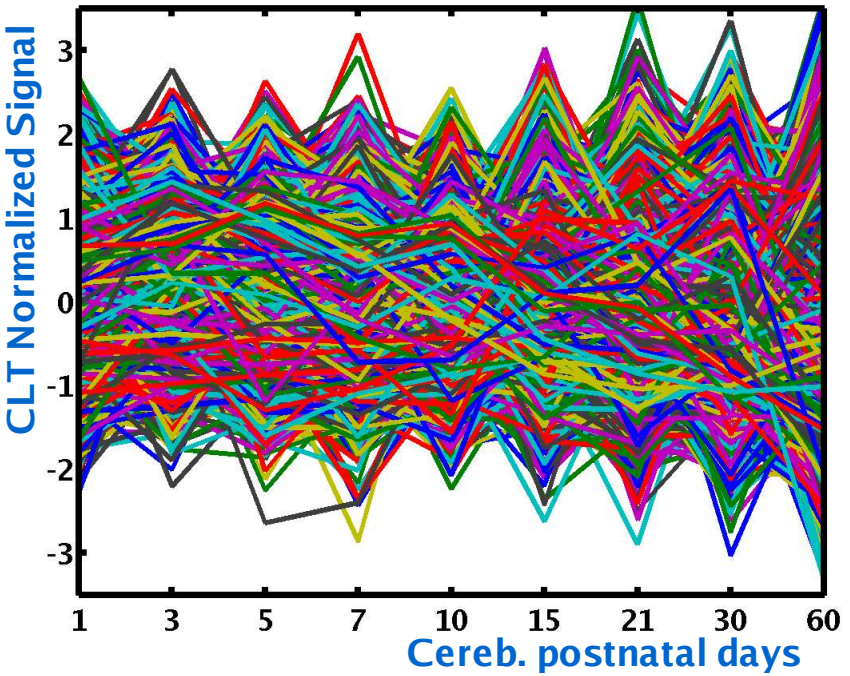


If we care about absolute intensities i.e., displacement, Euclidean distance between profiles





CLT* Normalized:
Avg 0
Var 1



** CLT = Central Limit Theorem scaling*

If we care about correlative "shape" i.e., first derivative, velocity, Pearson / Rank correlation between profiles

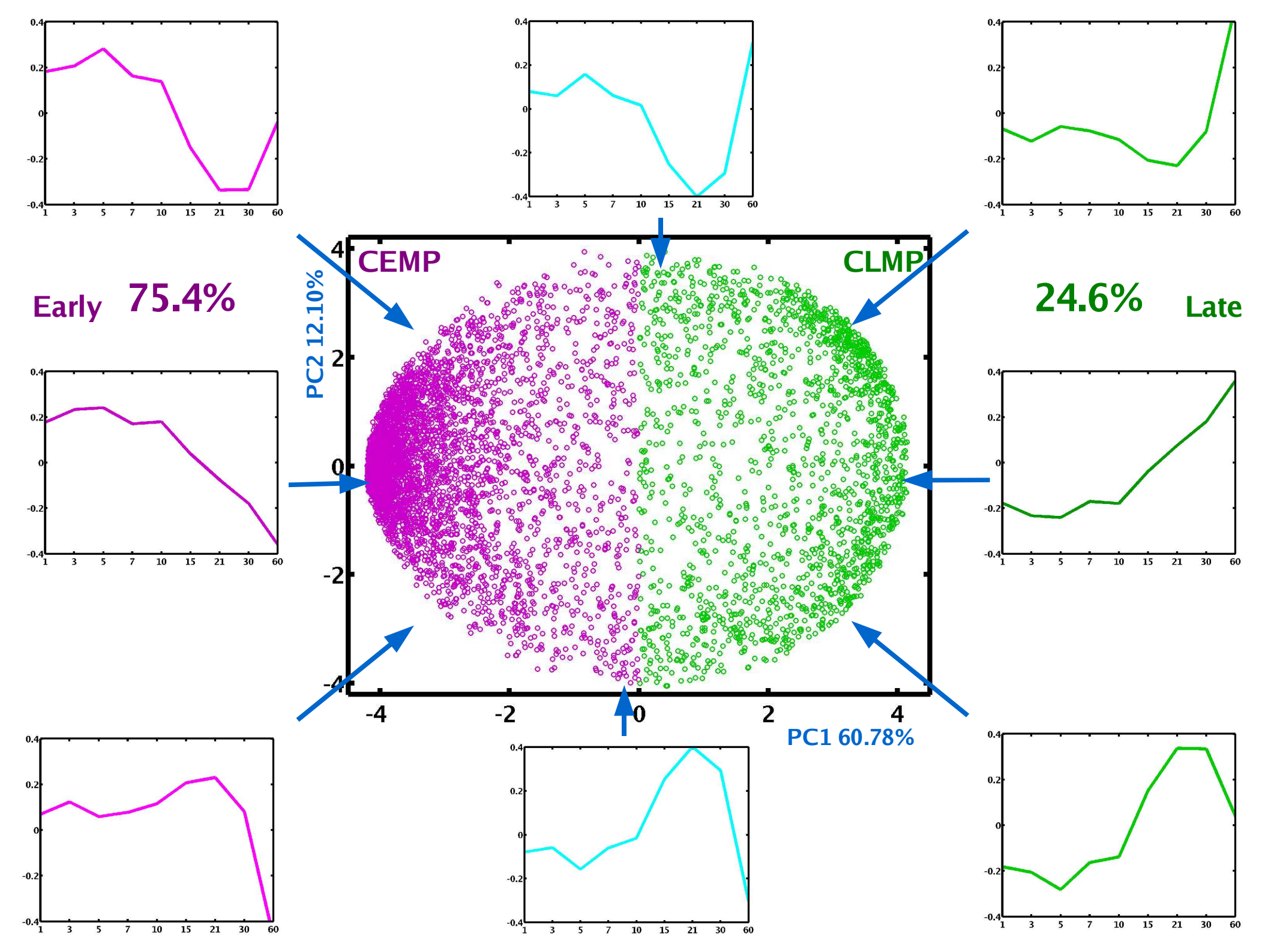
Modeling data: Visualizing time profiles of individual genes during mouse cereb. dev.

- ***Exercise***

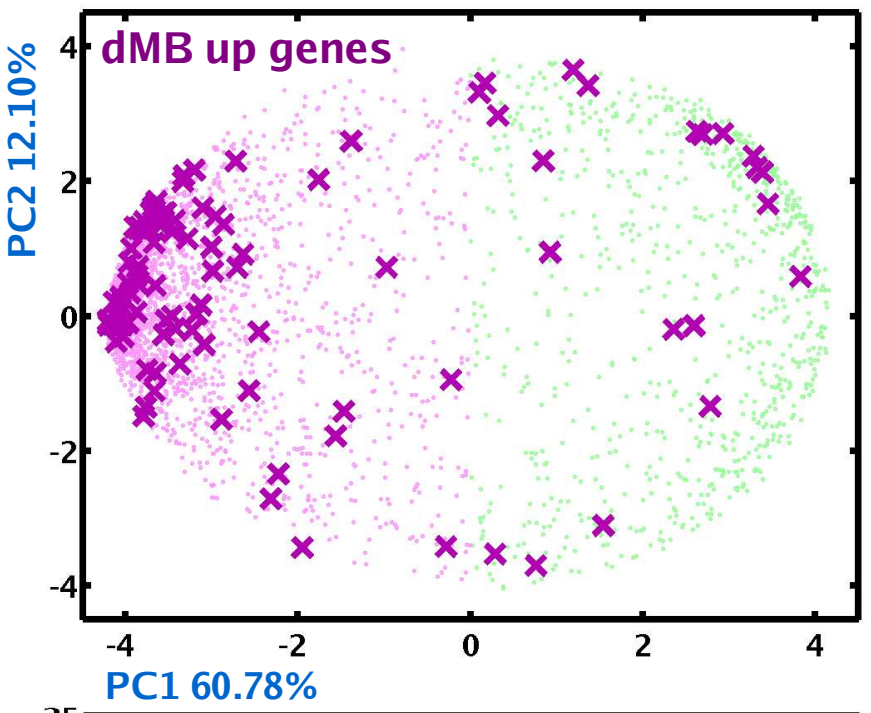
Given $\vec{x} = (x_1, x_2, \cdots, x_d) \in \mathbb{R}^d$, $d > 1$. Let $\mathbb{R}^d$ be equipped with the Euclidean metric and standard inner product $\langle \cdot, \cdot \rangle$. Show that $\vec{y} = (x_1 - \mu, x_2 - \mu, \cdots, x_d - \mu)$ is orthogonal to $\vec{1} = (1, 1, \cdots, 1)$, where $\mu_{\vec{x}} = \frac{1}{d} \sum_{j=1}^{d} x_j$. For $\sigma_{\vec{x}}^2 = \frac{1}{d-1} \sum_{j=1}^{d} (x_j - \mu)^2$, show that $\vec{y}/\sigma_{\vec{x}}$ lives on a hypersphere centered at $\vec{0}$ of radius $\sqrt{d-1}$ in $\mathbb{R}^d$.

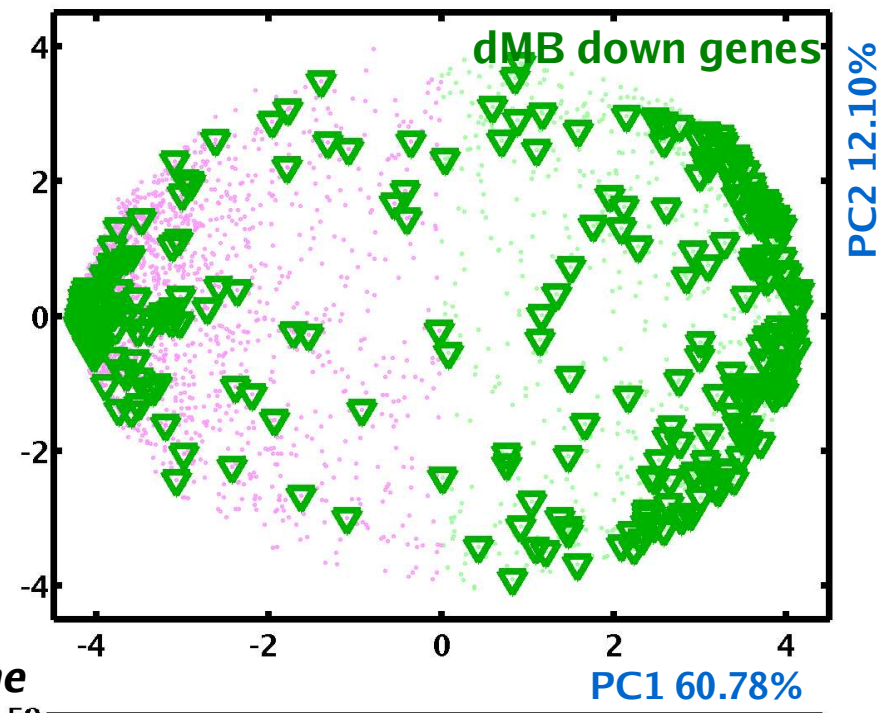
Let $\vec{u}$ and $\vec{v}$ have the properties $\mu_{\vec{u}} = \mu_{\vec{v}} = 0$ and $\sigma_{\vec{u}} = \sigma_{\vec{v}} = 1$ Show that $\rho(\vec{u}, \vec{v})$ – Pearson's correlation between $\vec{u}$ and $\vec{v}$ – is equivalent $\langle \vec{u}, \vec{v} \rangle$. Show that $e^2(\vec{u}, \vec{v}) = 2\,(1 - \rho(\vec{u}, \vec{v}))$, where $e(\vec{u}, \vec{v})$ denotes the Euclidean distance between $\vec{u}$ and $\vec{v}$.

- For now, we are interested in the "shapes" or first derivative of the RNA profiles of mouse cereb dev. CLT-normalize each individual RNA profile
- Apply *Principal Component Analysis / PCA* on mouse genes which are individually CLT-normalized across the cereb. dev. time points.

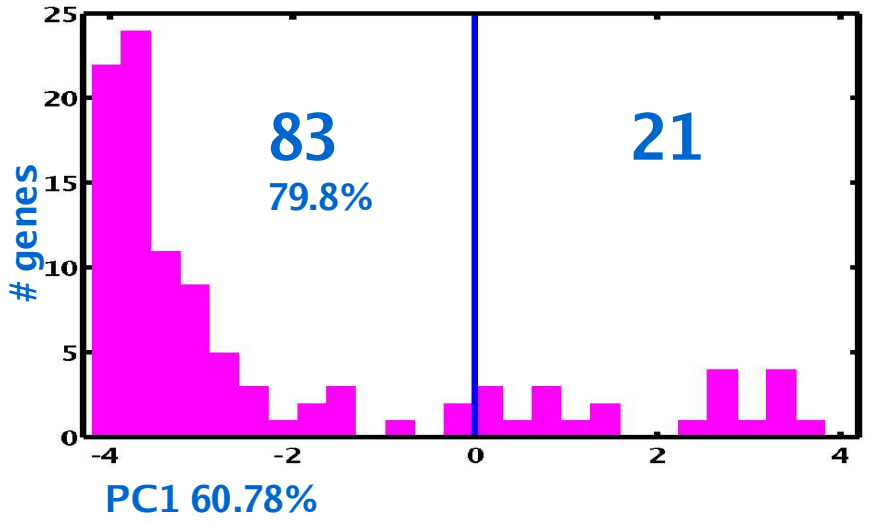

CEMP
CLMP
Early 75.4%
24.6% Late
PC2 12.10%
PC1 60.78%

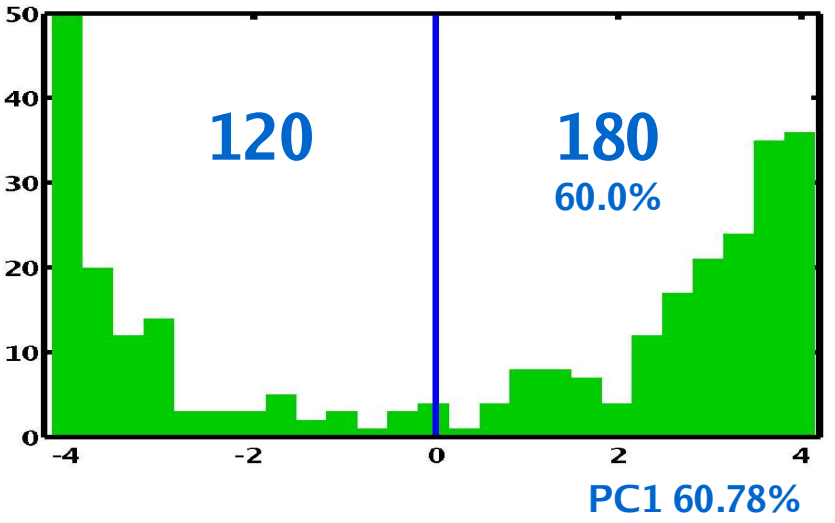
Modeling data: Human genes significantly Up, Down dMB/normal cereb. on mouse cereb. dev.





Molecular/ Gene-by-Gene Perspective

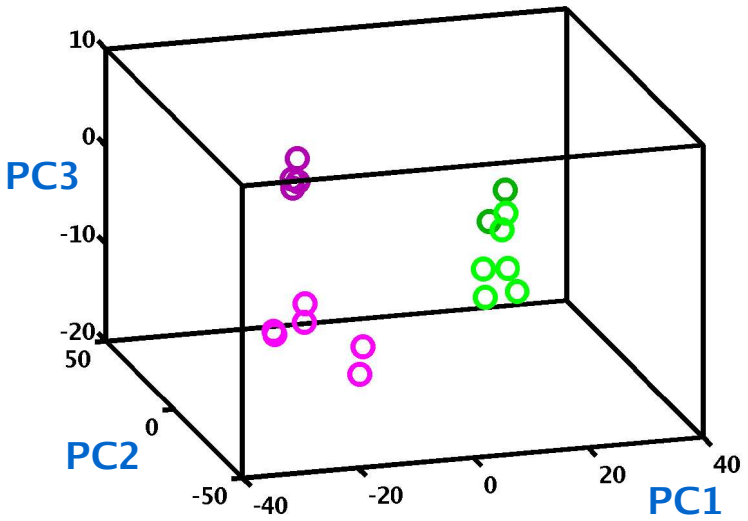


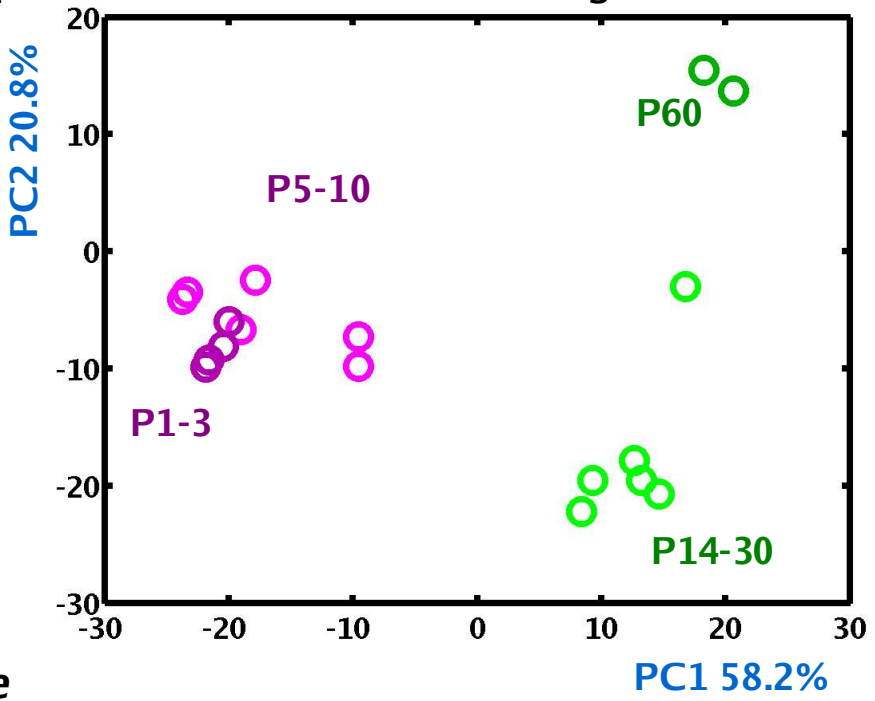


Are these distributions due to chance alone? Null hypothetical distribution = of all genes from previous slide (75.4% CEMP, 24.6% CLMP). Tests: Fisher exact, Chi square.

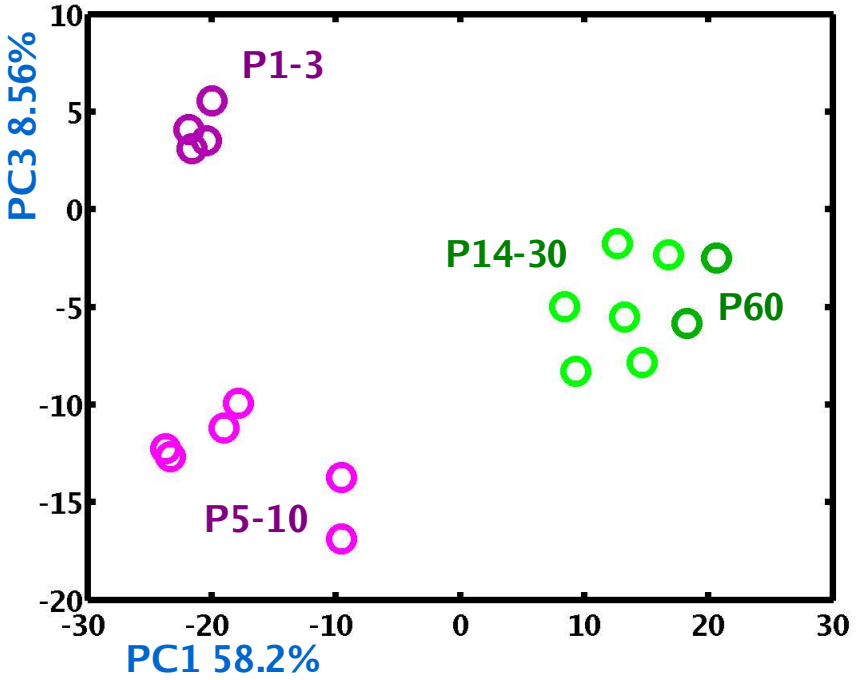
Modeling data: Visualizing mouse cereb. dev. genomic profile

Each mouse, human sample is a 3,477-vector. Each sample is CLT-normalized across all genes.



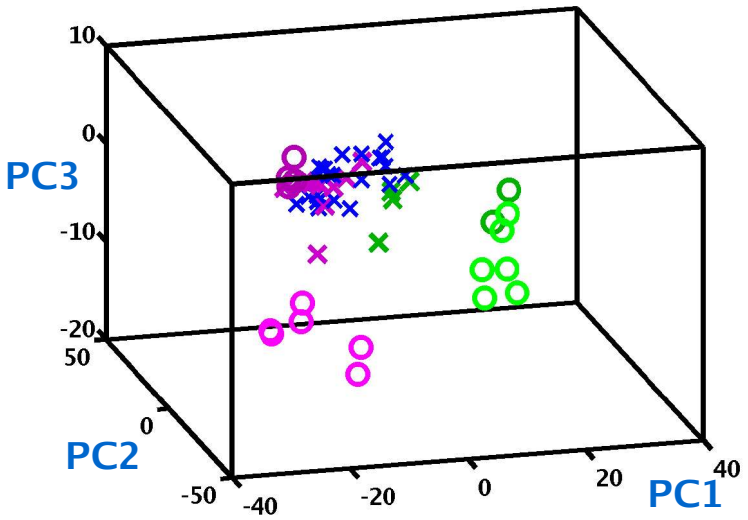


Global/ Genomic Perspective

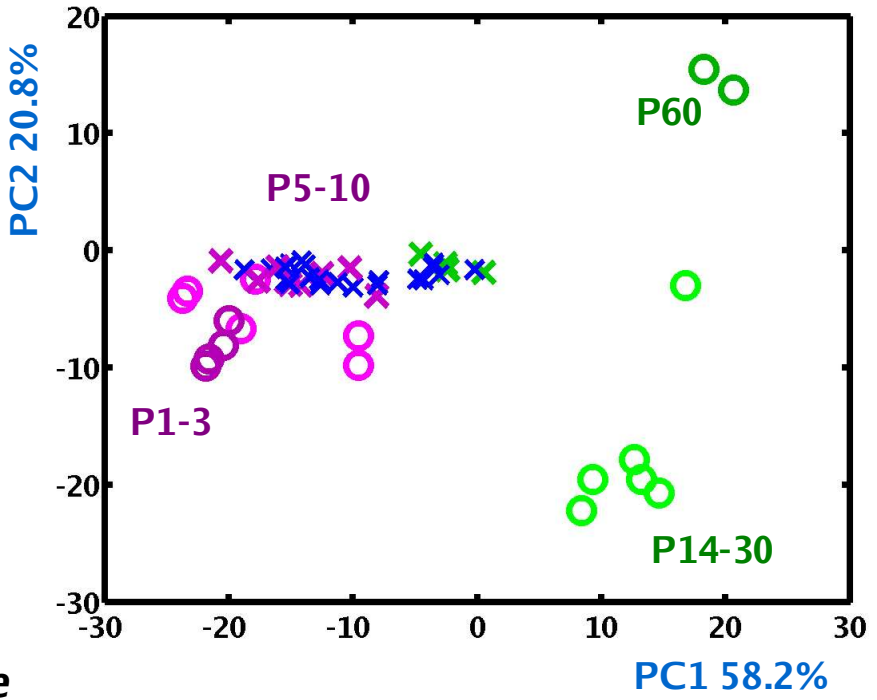


Let **M** be the 3,477 x 3,477 orthonormal matrix that maps each mouse vector into this new PC coordinates; and **x** a 3,477-vector in the original mouse coordinates / basis. Rewriting x in terms of new PC basis: **y = M'x**

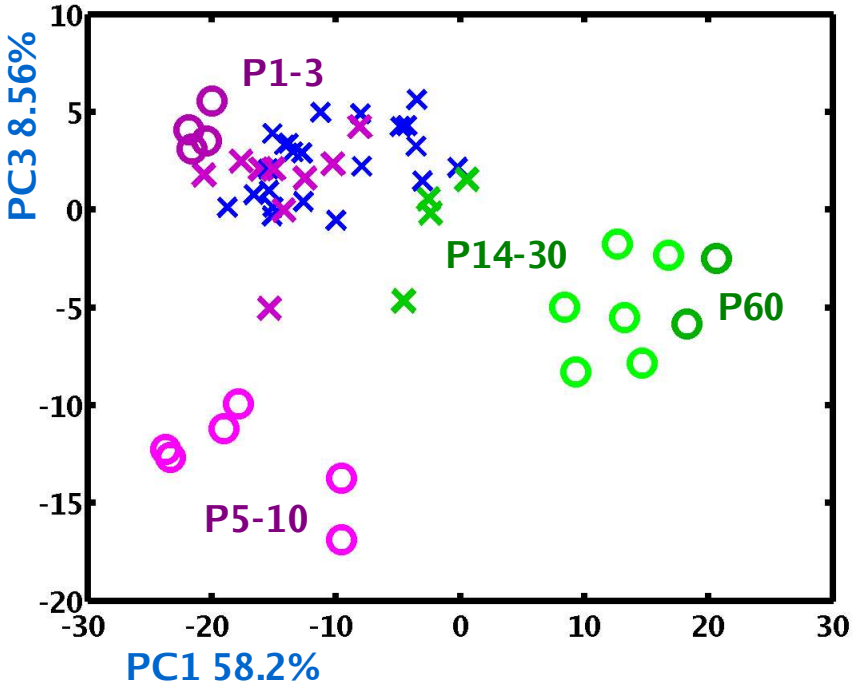
Modeling data: Human samples on mouse cereb. dev. genomic background



X normal
X dMB
X cMB

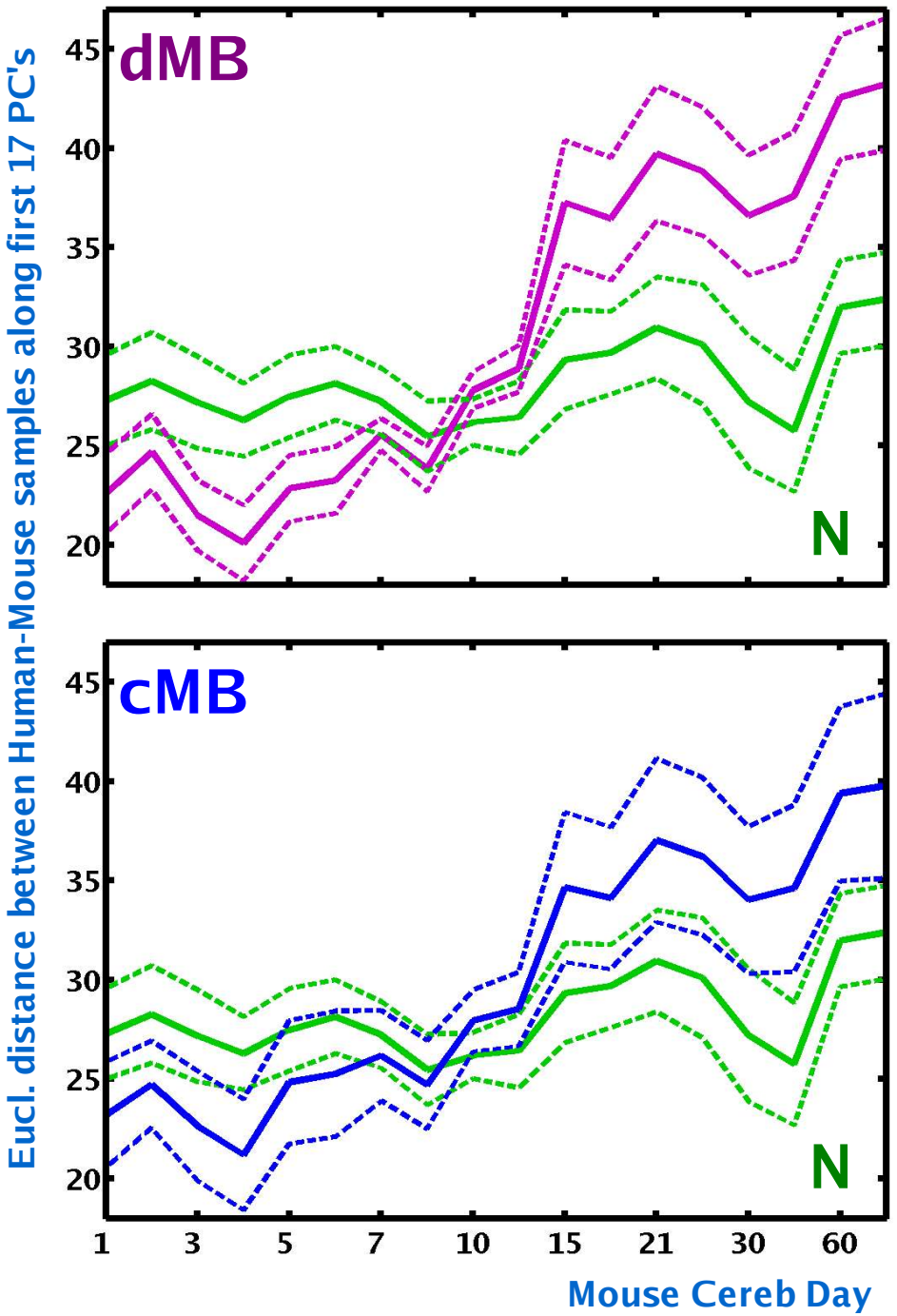


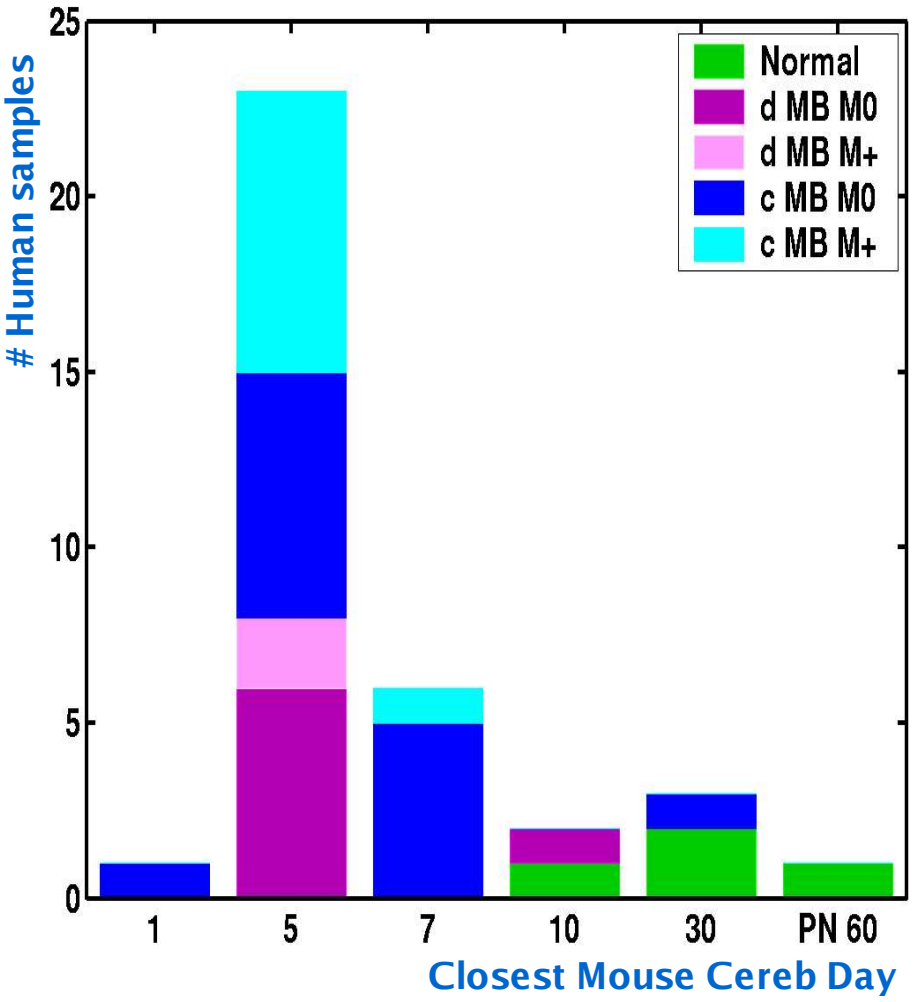
Global/ Genomic Perspective



Project every human sample into this mouse genomic PC space using the transformation overleaf where **x** is now a human 3,477-vector -- component-wise gene-homologous to mouse vector genes in original coordinates. **y = M'x**

Modeling data: Human samples against mouse cereb. dev. background





M0 non-metastatic
M+ metastatic

Global/ Genomic Perspective

Modeling data: Results

1 *Molecular / Gene-by-gene perspective:*

- Genes up-regulated in human MB tend to be ones more highly expressed in early mouse cereb. dev.
- Genes down-regulated in human MB tend to be ones more highly expressed in late mouse cereb. dev.

2 *Global / Genomic perspective:*

- Human MB's are more similar to mouse cereb. P1-10, esp. P3-5
- Human normal cereb are more similar to mouse cereb. P30-60
- Classic MB's have a more heterogeneous genomic profile than desmo. MB's
- Metastatic (stages 1-4) classic MB's tend to mouse cereb. P5 – i.e., we identified a putative genomic-scale developmental “counterpart” for MB metastasis

3 *Translating back to biological reality?*

Modeling data: Reality checks

1. Are the molecular / genomic segregation results peculiar to the brain?

   - ***Check 1:*** Non-brain system. Human normal lung and squamous cell lung carcinoma (SCL) on a mouse lung developmental background

2. What happens to segregations if a *"non-cognate"* developmental background was used?

   - ***Check 2:*** Human lung cancer on mouse CNS development
   - ***Check 3:*** Human CNS cancer on mouse lung development

3. Are cell-cycle related genes primarily driving a tumor's genomic similarity towards the early developmental stages?

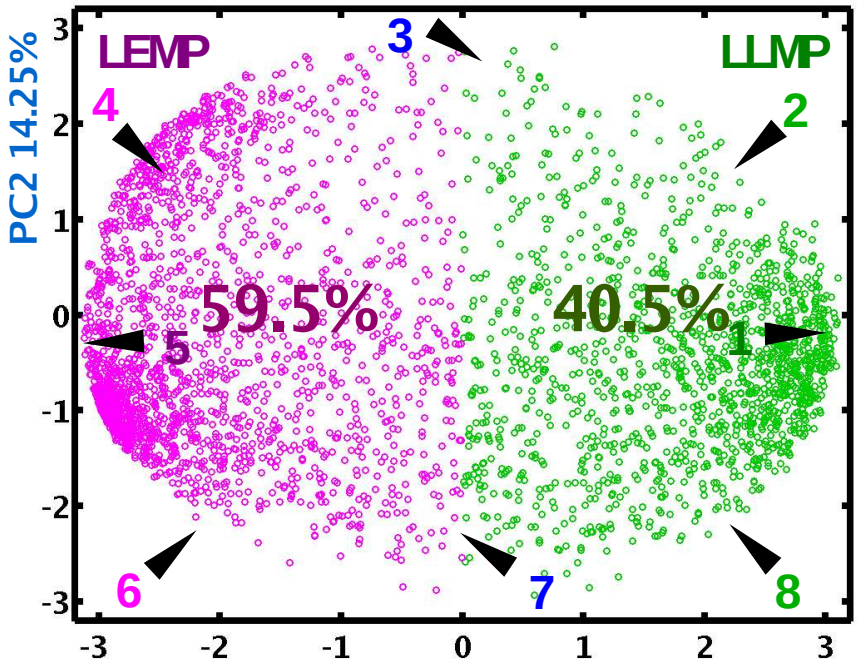
   - ***Check 4:*** Remove cell-cycle genes, re-run molecular and genomic analyses

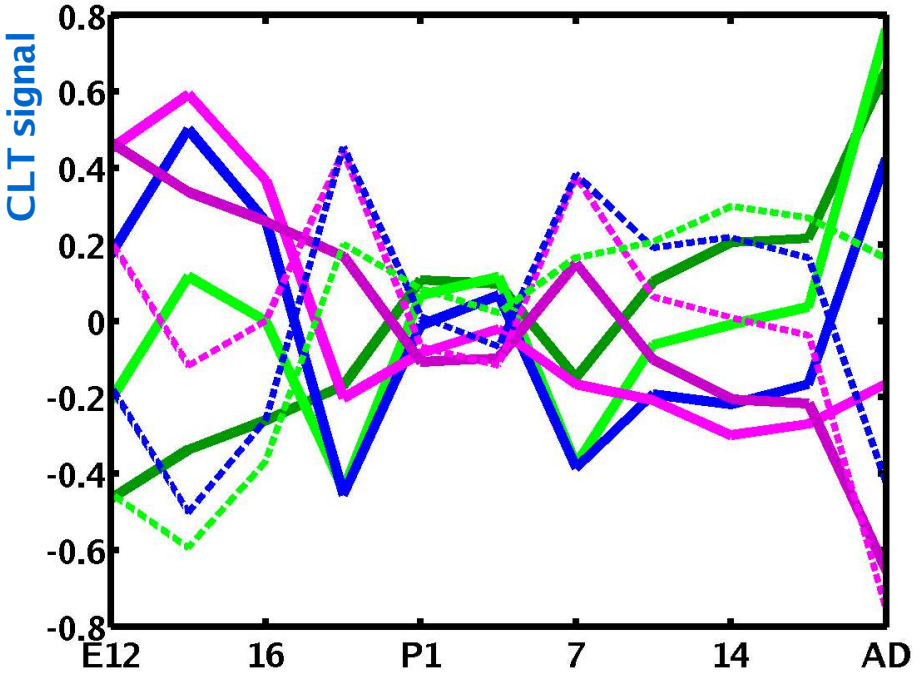
4. How are mouse models of MB related to mouse cereb. dev.?

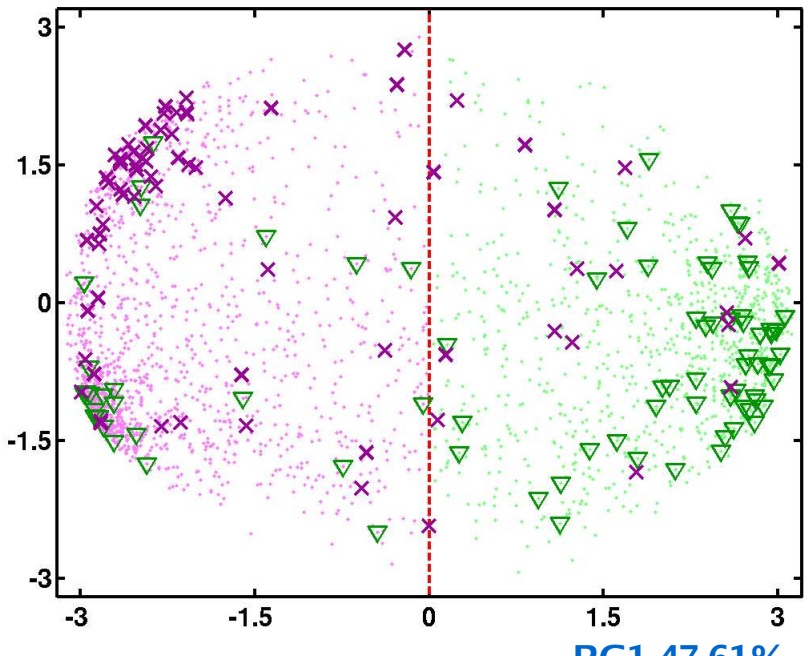
   - ***Check 5:*** Replace human MB's with *Ptch+/-* mouse MB's expression data
   - ***Check 6:*** How are human MB's related to *Ptch+/-* mouse MB's?

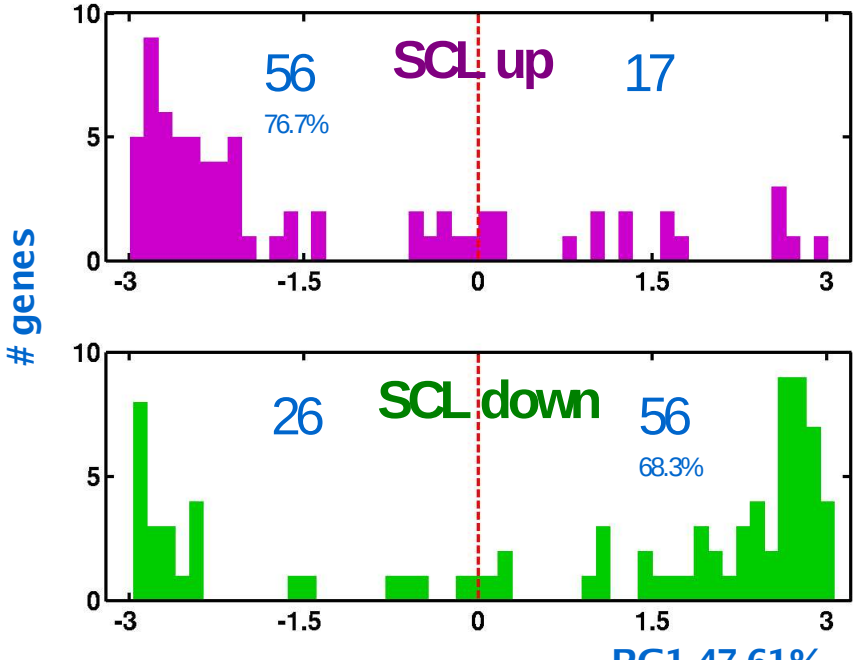
Modeling data: Reality Check 1 ~ what about non-brain system?

Human SCL, normal lung on mouse lung development





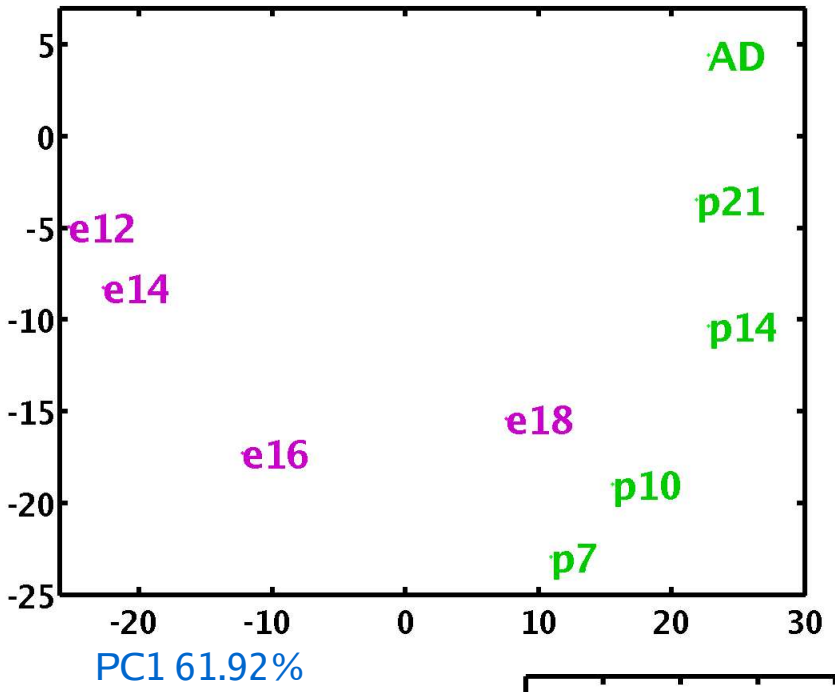


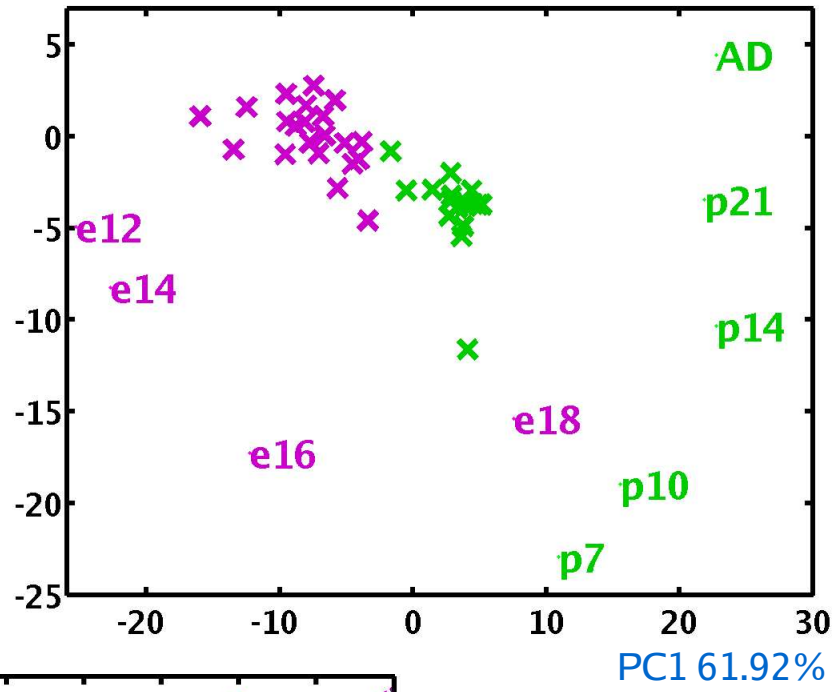


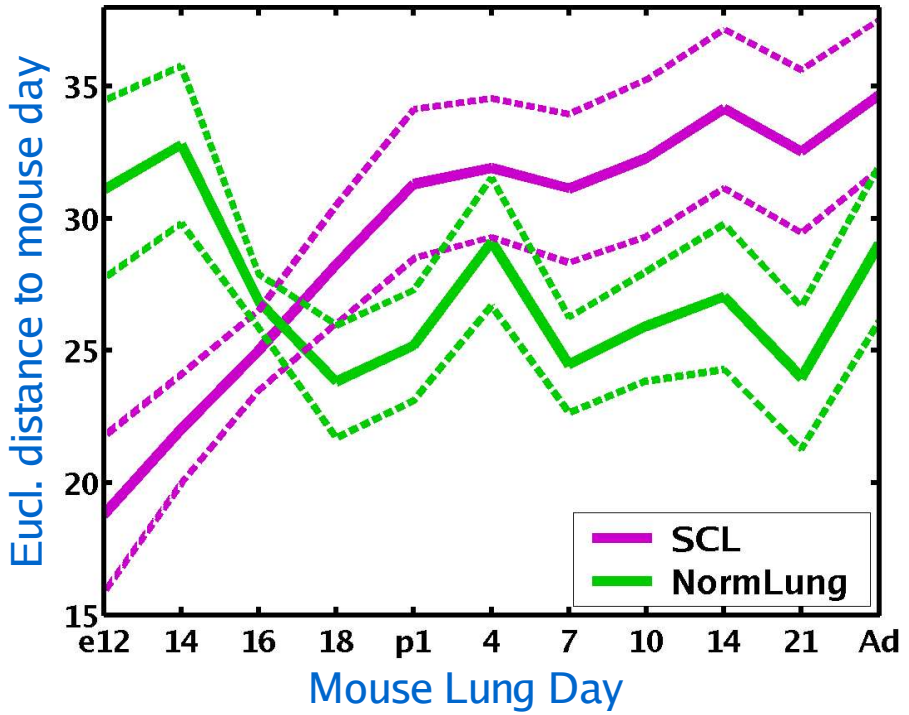
Molecular/ Gene-by-Gene Perspective

Modeling data: Reality Check 1 ~ what about non-brain system?

Human SCL, normal lung on mouse lung development



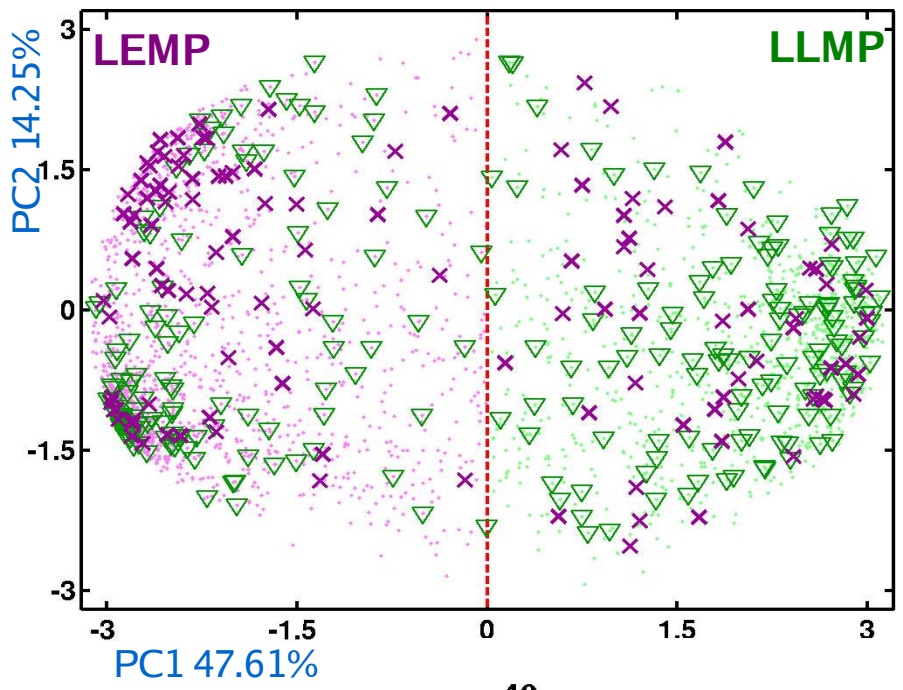


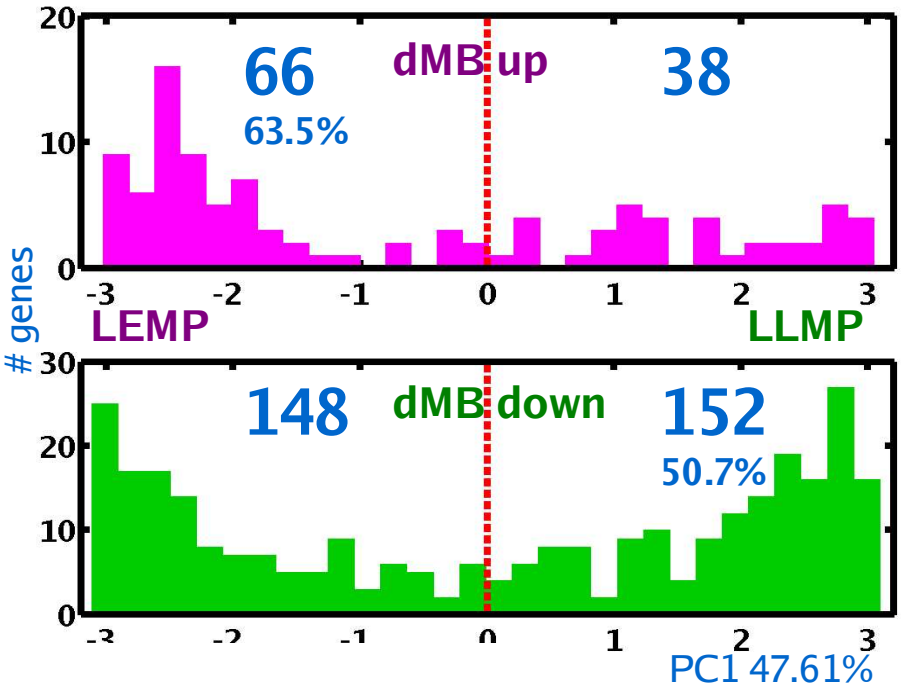


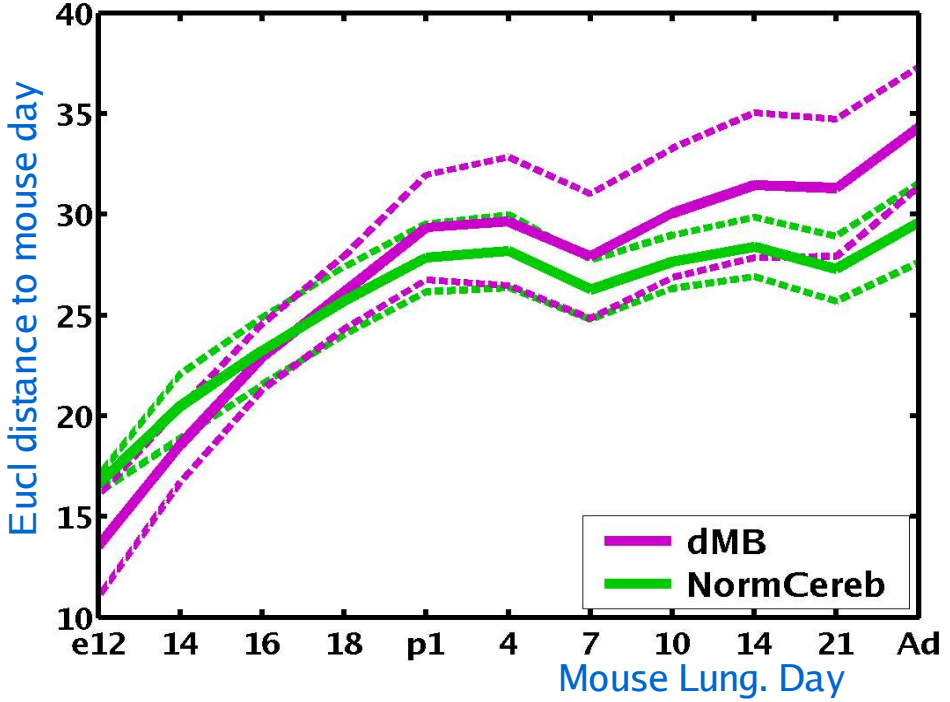
Global/ Genomic Perspective

Modeling data: Reality Checks 2, 3 ~ switching dev. backgrounds

Human dMB, normal cereb on mouse lung development

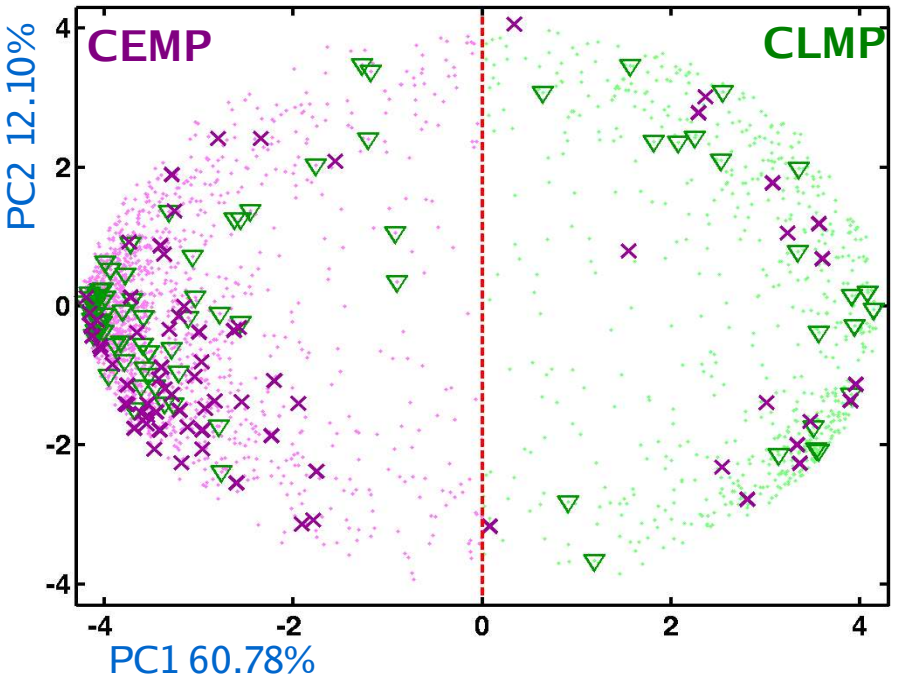


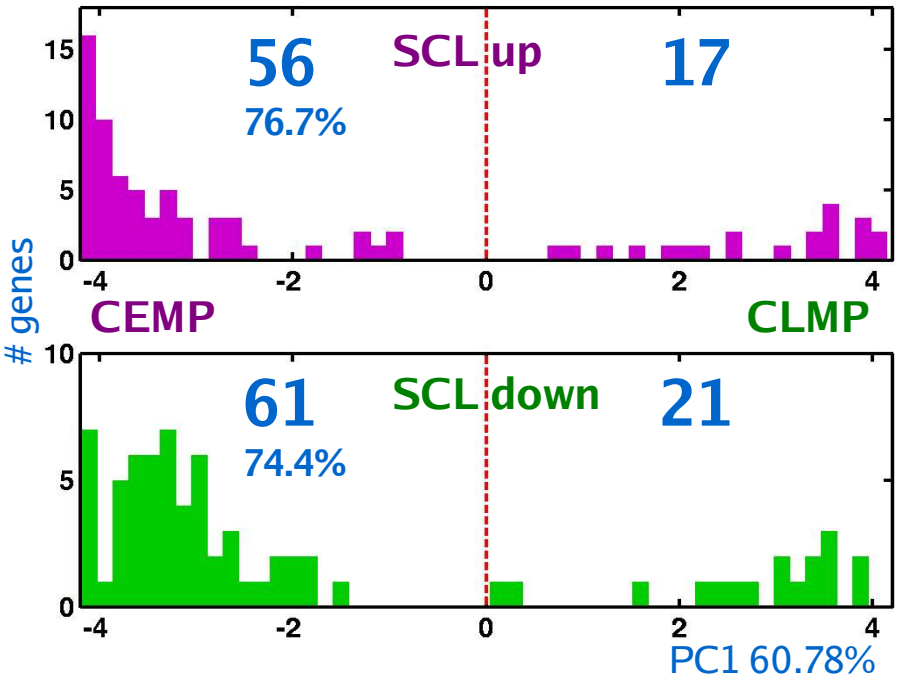


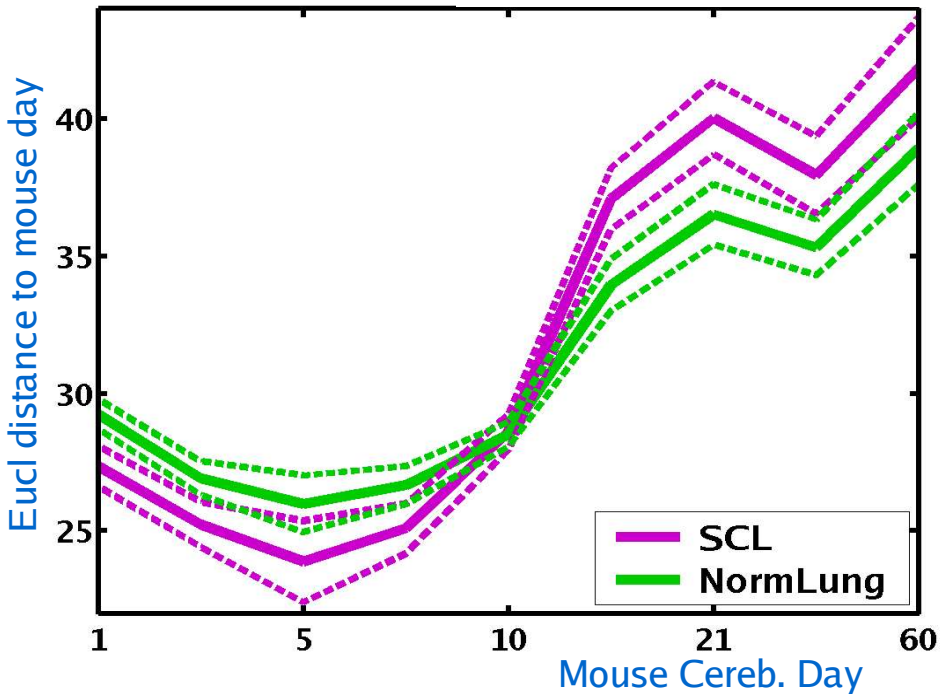


Modeling data: Reality Checks 2, 3 ~ switching dev. backgrounds

Human SCL, normal lung on mouse cereb development



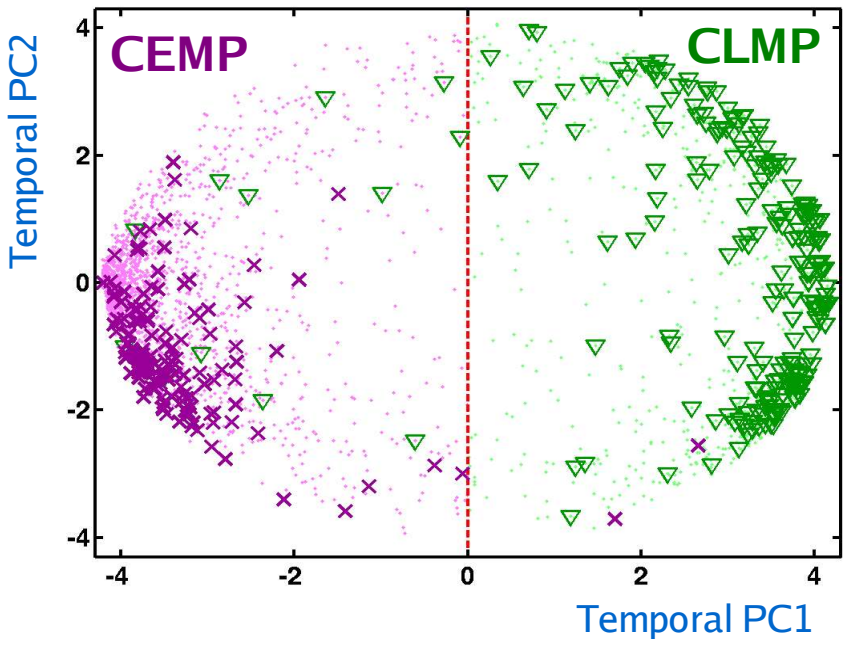


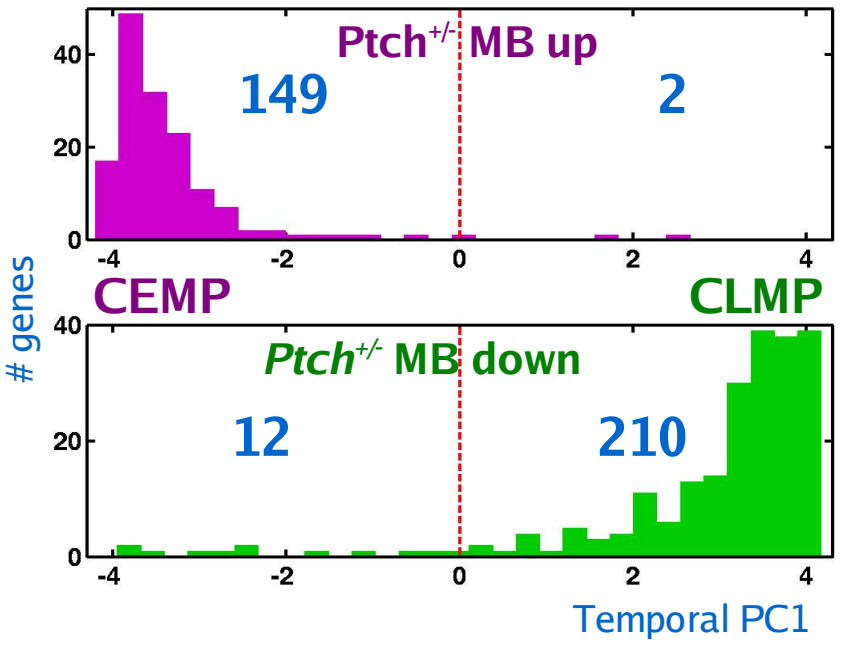


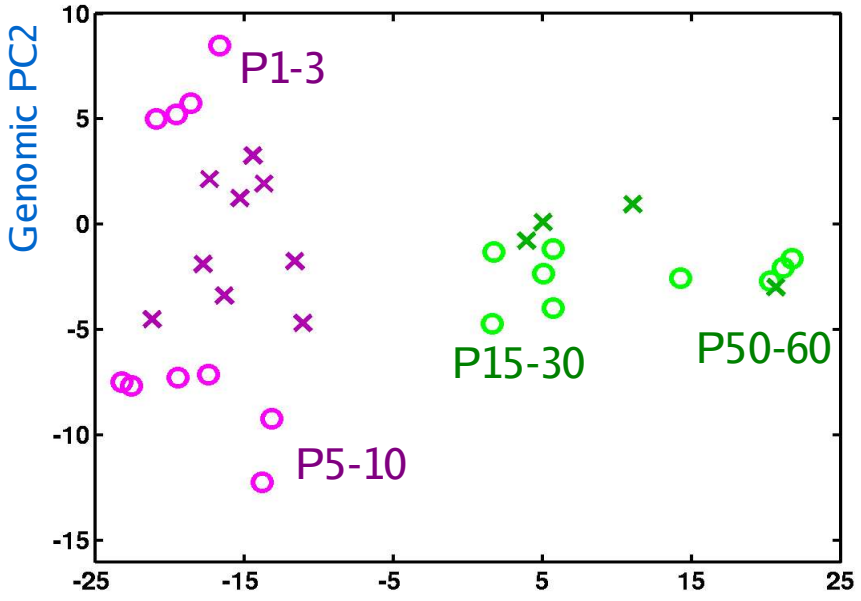
Modeling data: Reality Check 4 ~ Removing cell-cycle molecules

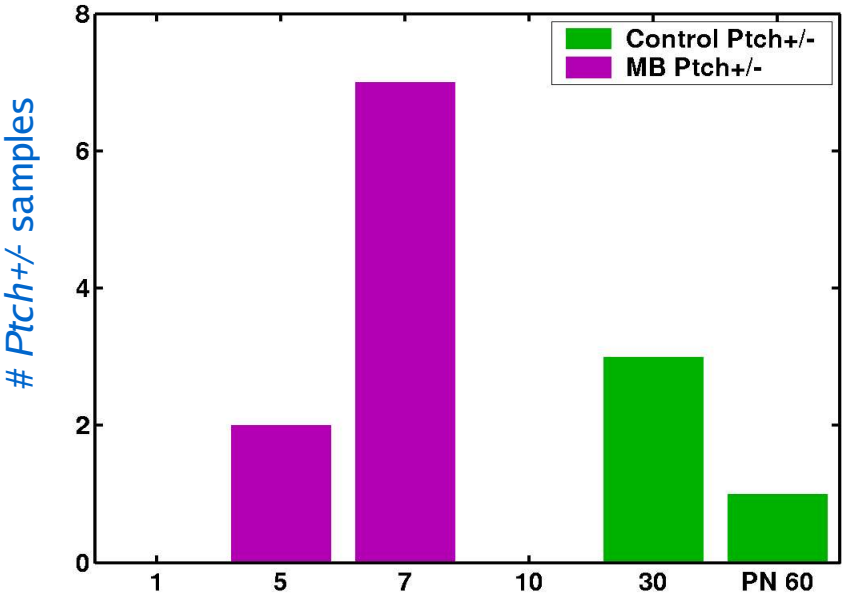
- ***Why cell cycle genes?*** A principal characteristic of cancer is abnormal / uncontrolled growth and proliferation of cells. Consequently, we expect to see cell-cycle related genes abundantly expressed in a tumor. Might this population of genes of a common function skew the genomic profile similarities vis-à-vis a developmental background?
  - ***Check 4:*** No, tumor-development genomic perspectives remain unchanged.

Modeling data: Reality Check 5 ~ *Ptch+/-* mouse MB's on mouse cereb dev.



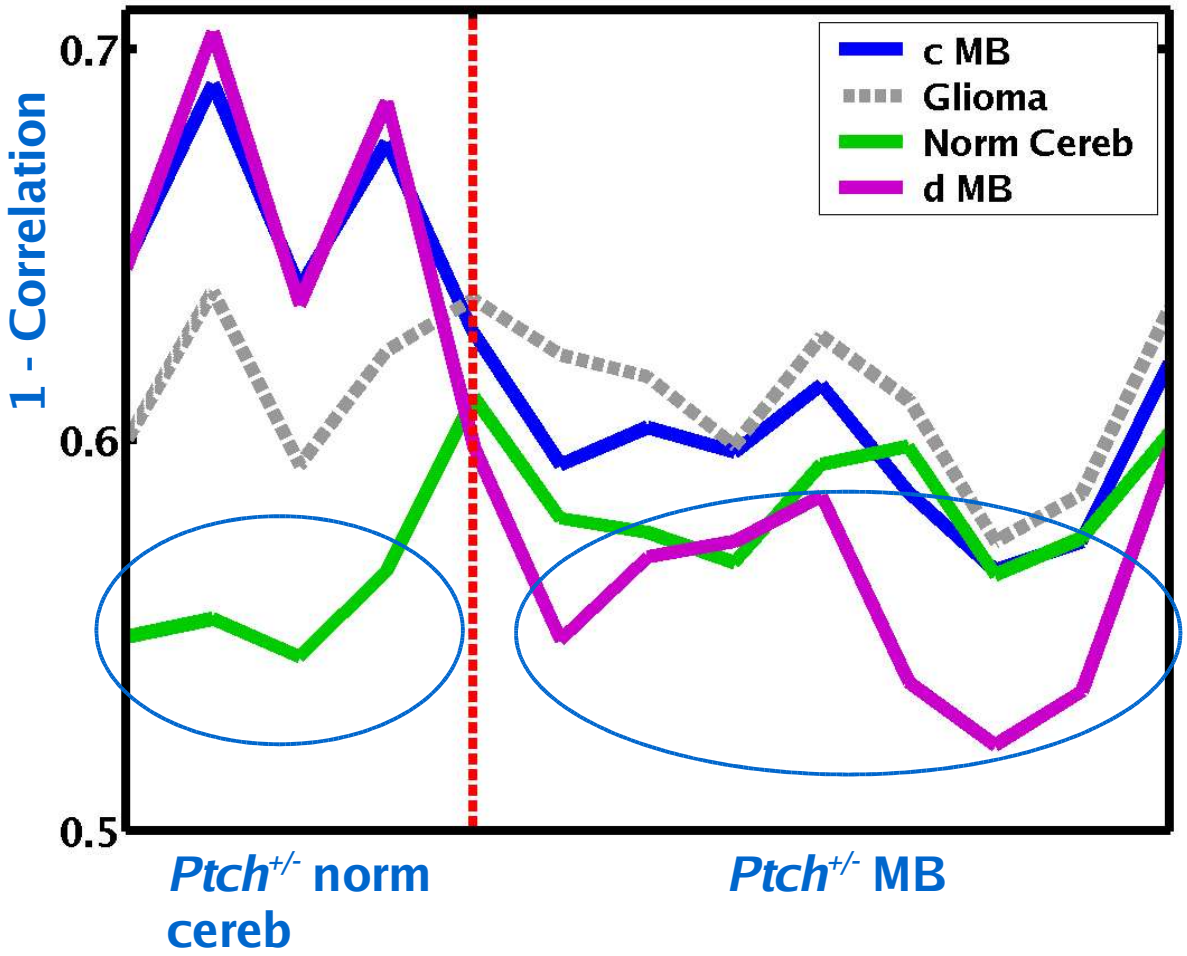




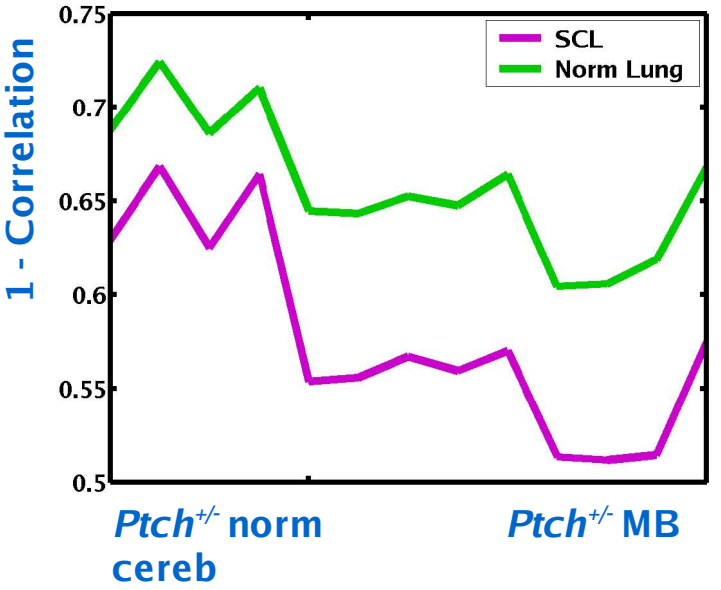


Modeling data: Reality Check 6 ~ Human MB's vs. *Ptch+/-* mouse MB's

Average human cereb sample genomic corr vs. each *Ptch+/-* mouse sample



Avg human lung sample genomic corr vs. each *Ptch+/-* mouse sample



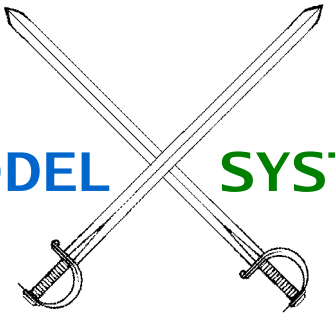
Human dMB's appear most genomically correlated ("similar") to *Ptch+/-* mouse MB's

Translating Model Outcome into Biological Reality, Thoughts

- *Specific stages of mouse cereb. dev. have been associated with MB and its clinical phenotype – metastasis – at a genomic scale.*
  - What does this mean for the biologist?
  - Does our notion of "similarity" coincide with the biologic notion of shared mechanisms / principal molecular underwriters for these 2 systems?
  - Are our experimental designs sufficient to establish biological "similarity"? No
- *Why study cancer on a developmental framework?*
  - Development is a dynamic, complex process which exercises – arguably – the full range of the genome / transcriptome in an organic way.
  - By definition, a pathologic condition = deviation from a "normal" reference frame. Current knowledge about cancer's intrinsic properties make development a natural reference system ... but which developing system?

Translating Model Outcome into Biological Reality, Thoughts

- *Modeling a biological system occurs at multiple levels*
  - Human disease > Mouse model > Data collection > Math model
  - Feature reduction + incorporation of comfounding factors (e.g., noise, heuristic assumptions) is inherent to each transition >
  - Assessing how effectively one level "represents" a previous level is important and non-trivial.

IF  THEN {make new model}

Epilogue

The discoveries that one can make with the microscope amounts to very little, for one sees with the mind's eye, and without the microscope, the real existence of all these little beings.

George-Louis Leclerc, Comte de Buffon 1707-1788

Some worthwhile references:

[1] *Philosophy of Mathematics and Natural Science*, H. Weyl, Princeton U. Press, 1949

[2] *The unreasonable effectiveness of mathematics in natural sciences*, E.P. Wigner, Comm. Pure & Applied Math. 13 (1), 1960

[3] *Applied Multivariate Statistical Analysis*, 4th edition, R.A. Johnson & D.W. Wichern, Prentice Hall, 1998

[4] *How to Solve It*, G. Polyà, Princeton U. Press 1957

[5] Numerous perspective & review articles on developmental biology in *Nature*, *Science* journals