

## Disease risk

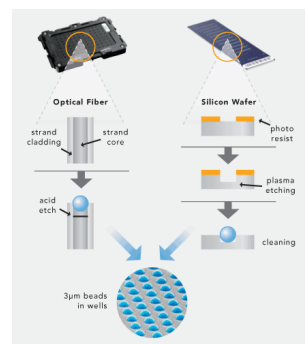
- Complex interplay between genetic and environmental risk
  - In familial (Mendelian) disorders high risk genetic variants with high penetrance, smaller input from environmental risks
  - In common diseases, low risk variants with very low penetrance + environmental risk

## Nanogenomic Technologies

- Whole genome genotyping
  - Microchips covering  $> 2.5 \times 10^6$  SNPs
  - Common Variants
- Next generation high throughput massively parallel sequencing
  - Rare and very rare (common) variants
  - Targeted, whole-exome and whole-genome sequencing

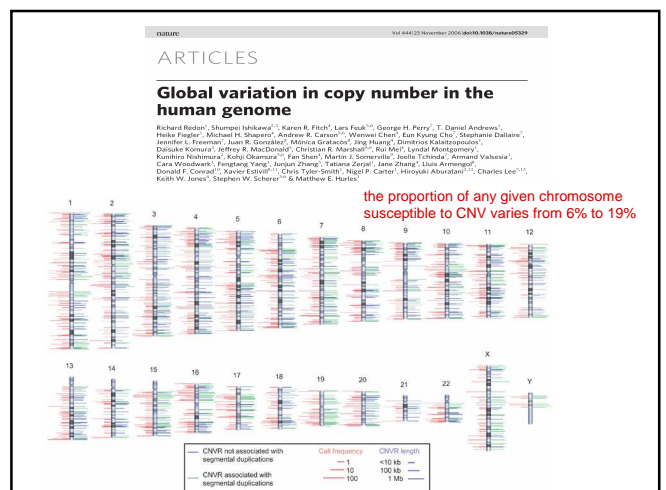
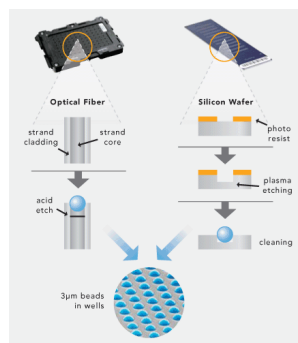
## Whole genome genotyping

- SNP chips ( $> 2,500,000$ ) for genome-wide:
  - Linkage analysis-homozygosity mapping
  - Copy number variation detection
  - Association analysis (GWAS)

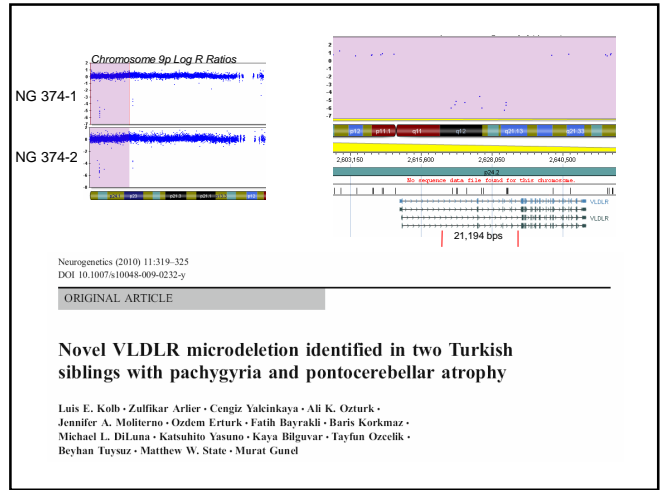
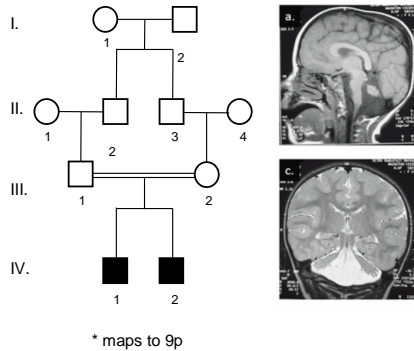


## Whole genome genotyping

- SNP chips ( $> 2,500,000$ ) for genome-wide:
  - Linkage analysis-homozygosity mapping
  - Copy number variation detection
  - Association analysis (GWAS)

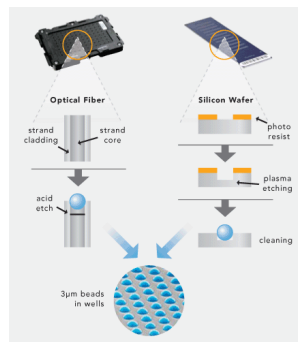


### Cerebellar Atrophy and Pachygyria\*-I



### Whole genome genotyping

- SNP chips (>2,500,000) for genome-wide:
  - Linkage analysis-homozygosity mapping
  - Copy number variation detection
  - Association analysis (GWAS)
    - Brain aneurysms



### Intracranial Aneurysm (IA)

- balloon-like dilations of cerebral arteries



### COMMON IA VARIANTS

### Common Variants in IA

- Based on genome-wide SNP genotyping with microchips
- Genome-wide Association Study (GWAS) to identify common variants that influence the occurrence and/or rupture of IA
  - based on the evidence for
    - high population prevalence (~2% by age 60)
    - substantial genetic contribution ( $\lambda_{\text{sub}}=4$ )
- a major strength of unbiased nature
  - surveying the entire genome without preconceptions regarding disease pathophysiology {WTCCC, 2007}

## IA GWAS

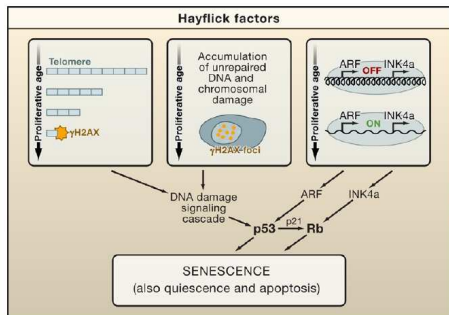
- a total of 20,072 subjects
  - 5,891 cases and 14,181 controls
  - discovery cohort of 2,780 cases and 12,515 controls
  - replication cohort of 3,111 cases and 1,666 controls
- SNP genotyping
  - analyses on 831,529 SNPs (directly genotyped and imputed)
- power to detect common variants (PFs  $\geq 5\%$ ) with OR of 1.25 and 1.20, respectively

## IA GWAS

| Locus    | SNP        | Position    | Genes          | Risk allele | Cohort      | P value               | $\log_{10}$ (Bayes) | PFA                   | Per-allele OR (95% CI) |
|----------|------------|-------------|----------------|-------------|-------------|-----------------------|---------------------|-----------------------|------------------------|
| 8q11.23  | rs10958409 | 55,489,644  | SOX17          | A           | Discovery   | $4.2 \times 10^{-7}$  | 4.64                | 0.8128                | 1.24 (1.14-1.35)       |
|          |            |             |                |             | Replication | 0.12                  | -0.11               |                       | 1.08 (0.98-1.20)       |
|          |            |             |                |             | Combined    | $9.0 \times 10^{-7}$  | 4.30                | 0.6685                | 1.17 (1.10-1.25)       |
| 8q12.1   | rs9298506  | 55,600,077  | SOX17          | A           | Discovery   | $1.2 \times 10^{-10}$ | 7.94                | 0.9999                | 1.33 (1.22-1.45)       |
|          |            |             |                |             | Replication | 0.0012                | 1.56                |                       | 1.21 (1.08-1.36)       |
|          |            |             |                |             | Combined    | $1.3 \times 10^{-12}$ | 9.55                | $1.0 \times 10^{-6}$  | 1.28 (1.20-1.38)       |
| 9p21.3   | rs1333040  | 22,073,400  | CDKN2A, CDKN2B | T           | Discovery   | $2.5 \times 10^{-16}$ | 13.41               | $1.0 \times 10^{-10}$ | 1.32 (1.24-1.41)       |
|          |            |             |                |             | Replication | $1.0 \times 10^{-7}$  | 5.18                |                       | 1.31 (1.19-1.45)       |
|          |            |             |                |             | Combined    | $1.5 \times 10^{-22}$ | 19.48               | $1.0 \times 10^{-16}$ | 1.32 (1.25-1.39)       |
| 10q24.32 | rs12413409 | 104,709,086 | CNM2           | G           | Discovery   | $7.9 \times 10^{-7}$  | 4.29                | 0.6621                | 1.38 (1.22-1.57)       |
|          |            |             |                |             | Replication | 0.00014               | 2.34                |                       | 1.23 (1.10-1.37)       |
|          |            |             |                |             | Combined    | $1.2 \times 10^{-9}$  | 7.00                | 0.9990                | 1.29 (1.19-1.40)       |
| 13q13.1  | rs9315204  | 32,591,837  | KL, STARD13    | T           | Discovery   | $3.3 \times 10^{-7}$  | 4.73                | 0.8443                | 1.21 (1.13-1.31)       |
|          |            |             |                |             | Replication | 0.0019                | 1.36                |                       | 1.18 (1.05-1.31)       |
|          |            |             |                |             | Combined    | $2.5 \times 10^{-9}$  | 6.72                | 0.9981                | 1.20 (1.13-1.28)       |
| 18q11.2  | rs11661542 | 18,477,693  | RBBP8          | C           | Discovery   | $5.6 \times 10^{-9}$  | 6.39                | 0.9959                | 1.21 (1.14-1.30)       |
|          |            |             |                |             | Replication | $4.5 \times 10^{-6}$  | 2.79                |                       | 1.22 (1.11-1.34)       |
|          |            |             |                |             | Combined    | $1.1 \times 10^{-12}$ | 9.92                | $1.0 \times 10^{-6}$  | 1.22 (1.15-1.28)       |

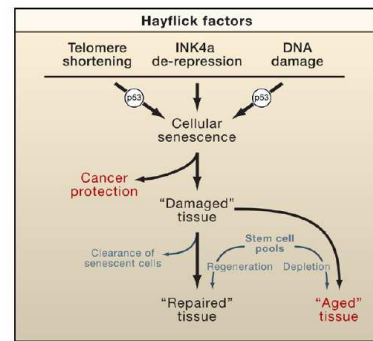
VOLUME 42 | NUMBER 5 | MAY 2010 | NATURE GENETICS

## Aging



Cellular Senescence in Cancer and Aging

Manuel Collado, Maria A. Blasco, and Manuel Serrano  
 Spanish National Cancer Research Center (CNIO), Madrid, Spain  
 DOI: 10.1016/j.cell.2007.07.009  
 Cell 130, July 27, 2007



Cellular Senescence in Cancer and Aging

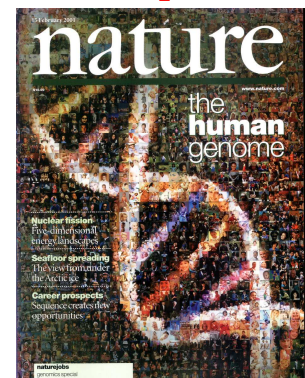
Manuel Collado, Maria A. Blasco, and Manuel Serrano  
 Spanish National Cancer Research Center (CNIO), Madrid, Spain  
 DOI: 10.1016/j.cell.2007.07.009  
 Cell 130, July 27, 2007

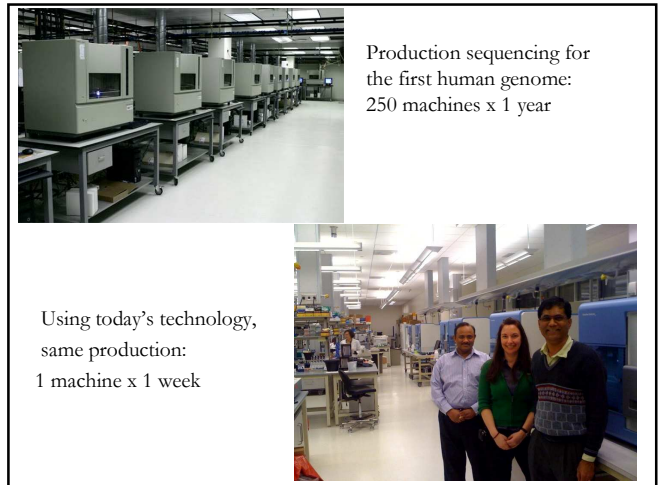
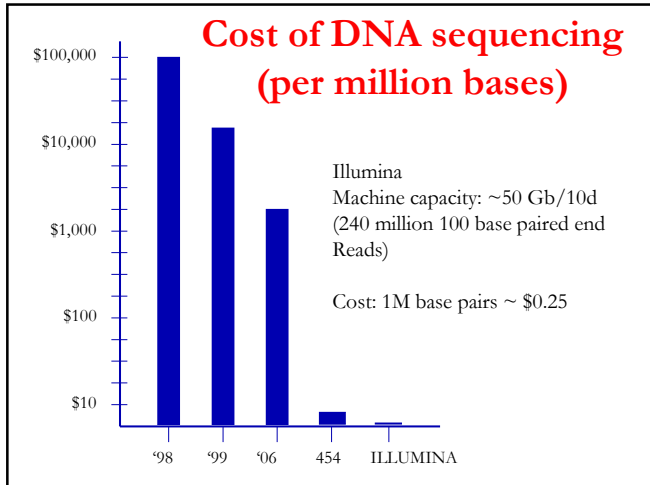
## Nanogenomic Technologies

- Whole genome genotyping
  - Microchips covering  $> 2.5 \times 10^6$  SNPs
  - Common Variants
- Next generation high throughput massively parallel sequencing
  - Rare and familial (very rare) variants
  - Targeted, whole-exome and whole-genome sequencing

## Human Genome Sequence

- Draft published 2001
- Complete finished sequence deposited in public databases April '03
- 3.1 billion bases/haploid genome
- ~ 20,000 genes





- ### Next Generation Re-sequencing
- New targeting strategies
    - enrich any targeted region 100-1,000 fold
  - Next generation parallel re-sequencing
    - Massive capacity
      - 1-2 x 10<sup>9</sup> (giga) base pair reads per run
  - Allows whole-exome sequencing
    - all coding regions (20,000 genes, ~196,000 exons) in the genome
  - Whole-genome sequencing
    - Significant reduction in cost
      - currently \$50,000/genome
      - expected to go as low as \$1,000

- ### DNA sequencing as a general-purpose tool in genetics
- Genome analysis
  - Transcriptome analysis
  - Regulatory site analysis
  - Disease gene discovery

### Genomes sequenced, 2000-2010

|                        |      |
|------------------------|------|
| • Viruses:             | 3029 |
| • Archaea:             | 172  |
| • Bacteria:            | 3252 |
| • Fungi:               | 236  |
| • Protists:            | 125  |
| • Plants:              | 120  |
| • Worms:               | 31   |
| • Insects:             | 53   |
| • Amphibians/reptiles: | 2    |
| • Fishes:              | 33   |
| • Mammals:             | 129  |

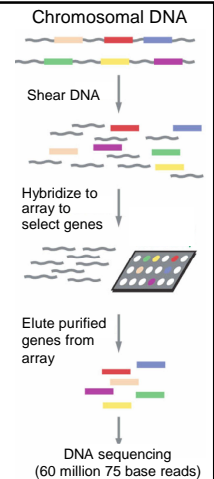
- ### Transcriptome analysis
- Northern blotting: 1977
  - Microarrays: 1995
  - Complete sequencing of transcriptomes: today
    - Can capture rare transcripts, splice variants, accurate quantitation
    - Measure changes with:
      - developmental stage
      - physiologic perturbation
      - mutation

## Gene Discovery in Disorders of the Brain and its Vasculature

- Common diseases
  - Common Variants
  - Rare Variants
    - Common variants explain only small fraction of risk of most common disease, < 10% for brain aneurysms
    - Current evidence consistent with rare alleles with moderate effect will explain large fraction of disease risk
      - Sequencing of case-control cohorts
      - Still prohibitively expensive
- Familial (Mendelian) genetics
  - Disease gene discovery
  - Exome sequencing

## Selective sequencing of all the genes in the human genome

- Genes comprise ~1% of the genome but harbor ~90% of the mutations with large effect
- Selective sequencing of all the genes can reduce cost of mutation discovery by 10-20 folds compared with whole genome sequencing
- Applicable to discovery of disease genes, therapeutic targets and clinical diagnosis

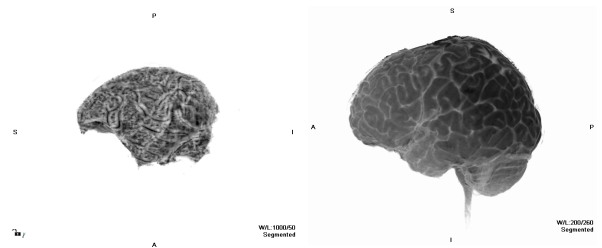


## Exome Sequencing

|  | Time          | Informatics pipeline  |
|--|---------------|---|
| • Make DNA from blood:   | 1 day         | Map each read to 'reference' human genome   |
| • Select genes from the genome and make library for sequencing:                          | 3 days        | ↓<br>Identify differences from reference sequence; filter for novel mutations             |
| • Sequence all the genes 50 times to ensure complete coverage (2.8 billion total bases): | 3.5 days      | ↓<br>Determine effect of DNA sequence change on encoded protein                           |
| • Run informatics pipeline to identify prioritized list of rare mutations:               | 0.5 days      | ↓<br>Determine likely effect of mutation on protein function from evolutionary constraint |
| <b>Total:</b>  | <b>9 days</b> |   |
| <b>Cost: ~\$3000</b>   |               |   |

## New opportunities in Mendelian genetics unmappable loci

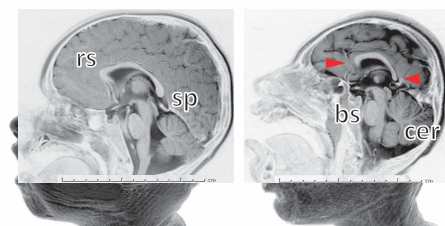
- Recessive mutations that severely impair brain development

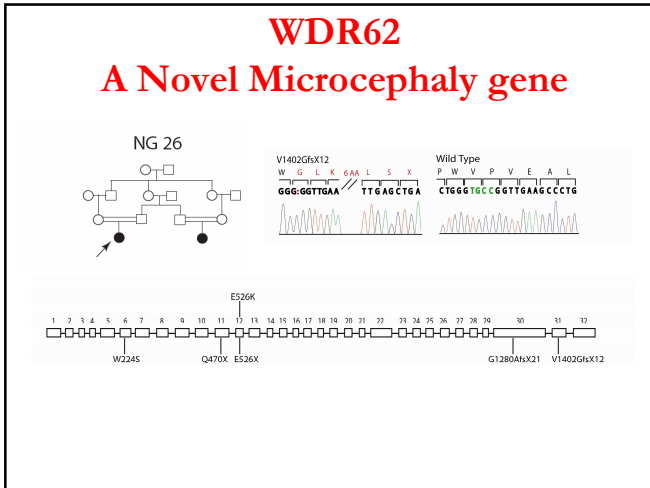


## Malformations of Cortical Development (MCD)

- discovery of genes hampered by:
  - phenotypic uncertainty
    - pleiotropy
  - significant locus and allelic heterogeneity
  - small consanguineous families
    - high early childhood mortality
    - typically single affected child
- new genomic technologies
  - genome vs exome sequencing

## Microcephaly Group I MCD- Proliferation

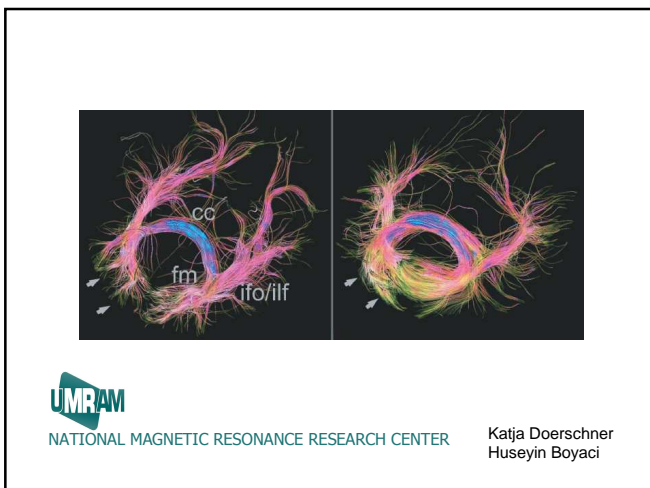
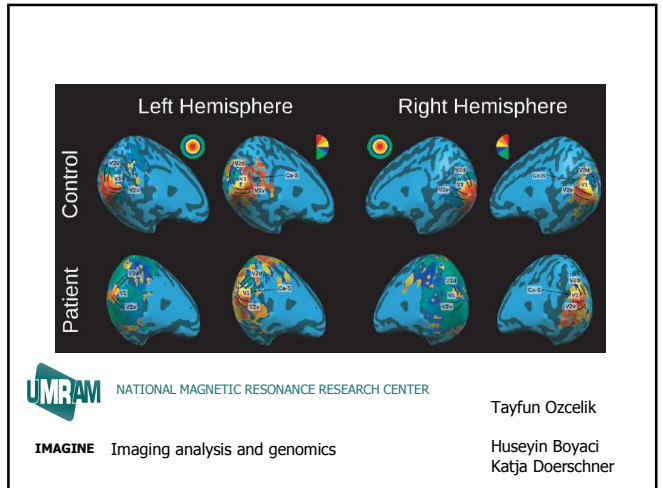
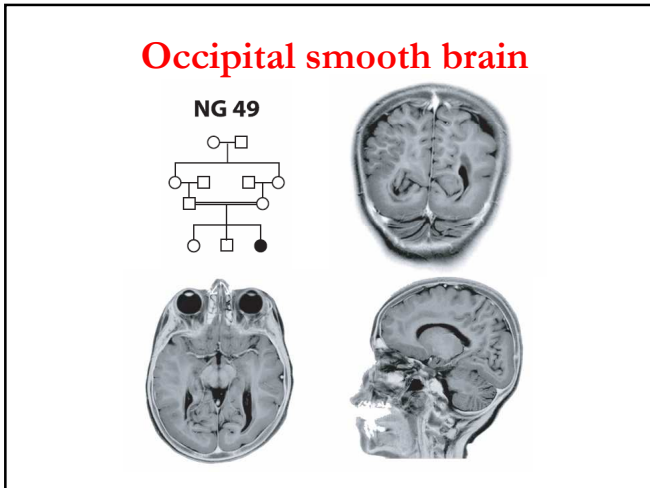




**Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations**

Kaya Bilgüvar<sup>1,2,3\*</sup>, Ali Kemal Öztürk<sup>1,2,3\*</sup>, Angeliki Louvi<sup>1,2,3</sup>, Kenneth Y. Kwan<sup>2,3</sup>, Murim Choi<sup>1</sup>, Burak Tatli<sup>1</sup>, Dilak Yılmazoğlu<sup>1</sup>, Beyhan Tüysüz<sup>1</sup>, Ahmet Okay Çalçıyan<sup>4</sup>, Sener Gükben<sup>5</sup>, Hande Kaymakçalan<sup>6</sup>, Tanyeri Barak<sup>1,2,3</sup>, Mehmet Bakırçöğür<sup>1,2,3</sup>, Katsuhito Yasuno<sup>1,2,3</sup>, Winson Ho<sup>1,2,3</sup>, Stephan Sanders<sup>1,11,12</sup>, Ying Zhu<sup>13</sup>, Sanem Yılmaz<sup>2</sup>, Alp Dinçer<sup>14</sup>, Michele H. Johnson<sup>14,15</sup>, Richard A. Bronen<sup>14</sup>, Naci Köçer<sup>16</sup>, Hüseyin Per<sup>17</sup>, Şirikant Manış<sup>18</sup>, Mehmet Necmettin Pamir<sup>19</sup>, Cengiz Yalçınkaya<sup>20</sup>, Selar Kamandany<sup>21</sup>, Meral Topcu<sup>22</sup>, Meral Özmen<sup>23</sup>, Nenad Sestan<sup>14</sup>, Richard P. Lifton<sup>1,2,3</sup>, Matthew W. State<sup>1,11,12</sup> & Murat Günel<sup>1,2,3</sup>

<sup>1</sup>Department of Neurosurgery, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>2</sup>Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>3</sup>Department of Genetics, Center for Human Genetics and Genomics and Program in Neurogenetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>4</sup>Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>5</sup>Division of Neurology, Department of Pediatrics, Istanbul University Istanbul Medical Faculty, Istanbul 34093, Turkey. <sup>6</sup>Division of Neurology, Department of Pediatrics, Hacettepe University School of Medicine, Sıhhiye, Ankara 06100, Turkey. <sup>7</sup>Division of Genetics, Department of Pediatrics, Istanbul University Cerrahpaşa Faculty of Medicine, Istanbul 34098, Turkey. <sup>8</sup>Department of Medical Genetics, Kayseri Education and Research Hospital, Kayseri 38010, Turkey. <sup>9</sup>Division of Neurology, Department of Pediatrics, Ege University Faculty of Medicine, Izmir 35100, Turkey. <sup>10</sup>Faculty of Arts and Sciences, Balıkesir University, Istanbul 34533, Turkey. <sup>11</sup>Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>12</sup>Child Study Center, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>13</sup>Department of Radiology, Acibadem University School of Medicine, Istanbul 34742, Turkey. <sup>14</sup>Department of Radiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>15</sup>Department of Otolaryngology, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>16</sup>Department of Radiology, Istanbul University Cerrahpaşa Faculty of Medicine, Istanbul 34098, Turkey. <sup>17</sup>Department of Radiology, Istanbul University School of Medicine, Istanbul 34098, Turkey. <sup>18</sup>Department of Radiology, Erciyes University School of Medicine, Kayseri 38039, Turkey. <sup>19</sup>Yale Center for Genome Analysis, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>20</sup>Department of Neurosurgery, Acibadem University School of Medicine, Istanbul 34742, Turkey. <sup>21</sup>Division of Child Neurology, Department of Neurology, Istanbul University Cerrahpaşa Faculty of Medicine, Istanbul 34098, Turkey. <sup>22</sup>Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>23</sup>These authors contributed equally to this work.



## Impact of Nanogenomic Technologies

- application of next generation genomic technologies have the power to identify disease loci in settings in which traditional methods have proved challenging:
  - locus and allelic heterogeneity
  - small pedigrees that cannot support independent statistical evidence for linkage
  - phenotypic uncertainty or pleiotropy
- defining genetic risk due to very-rare, rare or common variants
  - downstream analysis
- ultimately allowing for pre-clinical diagnosis, risk prediction and new therapies

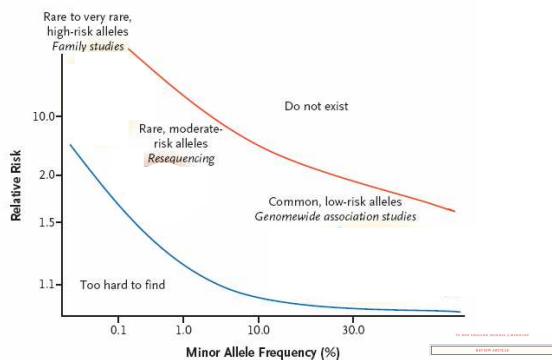
## Economic Impact

- Faster gene discovery
  - Causal relationship between molecules and diseases
  - New drug targets
  - Rational drug design
- Application to clinical medicine- Personalized medicine
  - People living longer-disease prediction
  - Somatic mutations in primary, metastatic and resistant cancers
  - Individualized medications for common diseases-hypertension, asthma, cardiovascular disease...
  - IVF, infectious diseases, newborn, etc
- The future...

## Acknowledgements

- Richard Lifton
- Matthew State
- Angeliki Louvi
- Shrikant Mane
- Katsuhito Yasuno
- Kaya Bilguvar
- Ali Ozturk
- Mehmet Bakircioglu
- Tanyeri Barak
- Winson Ho
- Kenneth Kwan
- YCGA Staff
- Andrea Chamberlain
- Collaborators throughout the world...
- Nenad Sestan
- Antonio Giraldez
- Mustafa Khokha
- Yale School of Medicine
- NIH (NINDS)

## Genetic Risk



## Applications of genetics to neuroscience

- Genetic tests with high predictive value for diagnosis and therapeutic response
  - Who should we sequence?
  - When should this be done?
  - How do we deal with incomplete understanding?
  - How do we communicate results?
  - Implications for education of health care professionals, patients, health and social policy

## Capture and Re-sequencing

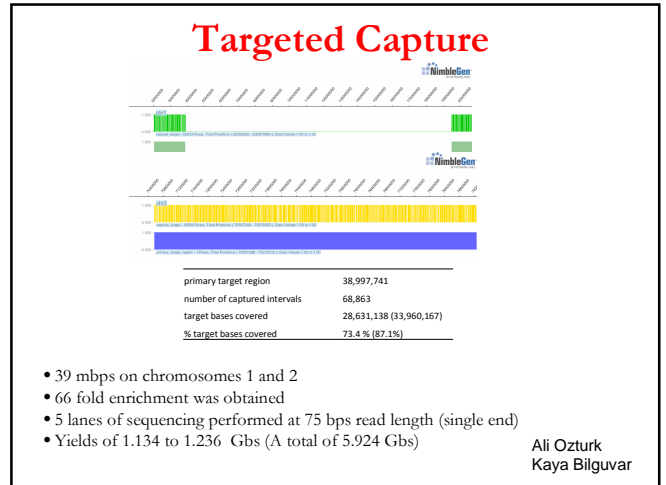
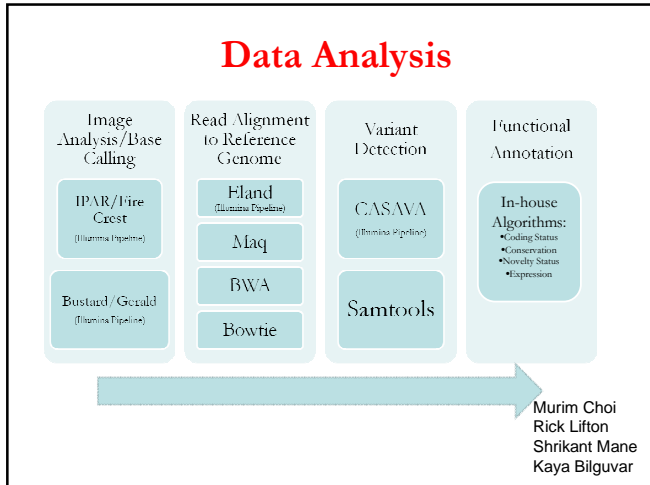
- Large linkage interval
  - >200 genes
- NimbleGen Sequence Capture 2.1M custom arrays to capture the entire affected-only linkage interval
- Genome Analyzer II by Solexa (Illumina)
  - 75 bp single read sequencing

Shrikant Mane  
 Rick Lifton

## Data Generation

- Current experimental yields from single read sequencing :
  - 8 lanes per run
  - ~ 130,000 clusters and ~17,500,000 reads per lane
  - ~ 1.2 Giga bases per lane (now 2.0 Giga bases per lane)

| Lane Info |                     | Tile Mean +/- SD for Lane |                 |                    |                                  |                |                |                      |                   |
|-----------|---------------------|---------------------------|-----------------|--------------------|----------------------------------|----------------|----------------|----------------------|-------------------|
| Lane      | Lane Yield (kbases) | Clusters (raw)            | Clusters (PF)   | 1st Cycle Int (PF) | % intensity after 20 cycles (PF) | % PF Clusters  | % Align (PF)   | Alignment Score (PF) | % Error Rate (PF) |
| 3         | 1134080             | 139337 +/- 5057           | 124350 +/- 9414 | 304 +/- 10         | 76.23 +/- 5.70                   | 89.25 +/- 6.25 | 88.65 +/- 1.98 | 263.69 +/- 7.39      | 0.50 +/- 0.13     |
| 5         | 1236775             | 153335 +/- 3957           | 135611 +/- 4544 | 292 +/- 7          | 75.13 +/- 1.45                   | 88.44 +/- 1.74 | 88.84 +/- 0.19 | 262.53 +/- 1.94      | 0.50 +/- 0.03     |
| 6         | 1206980             | 149432 +/- 3924           | 132344 +/- 4824 | 287 +/- 6          | 76.33 +/- 1.48                   | 88.55 +/- 1.66 | 88.86 +/- 0.17 | 264.04 +/- 1.73      | 0.49 +/- 0.03     |
| 7         | 1193888             | 147210 +/- 3562           | 130908 +/- 3115 | 279 +/- 7          | 76.82 +/- 1.79                   | 88.93 +/- 1.15 | 88.93 +/- 0.12 | 265.15 +/- 1.67      | 0.47 +/- 0.02     |
| 8         | 1152607             | 145319 +/- 5868           | 126382 +/- 5490 | 231 +/- 6          | 75.76 +/- 1.66                   | 86.99 +/- 2.31 | 88.70 +/- 0.24 | 260.71 +/- 4.41      | 0.50 +/- 0.03     |



## Coverage

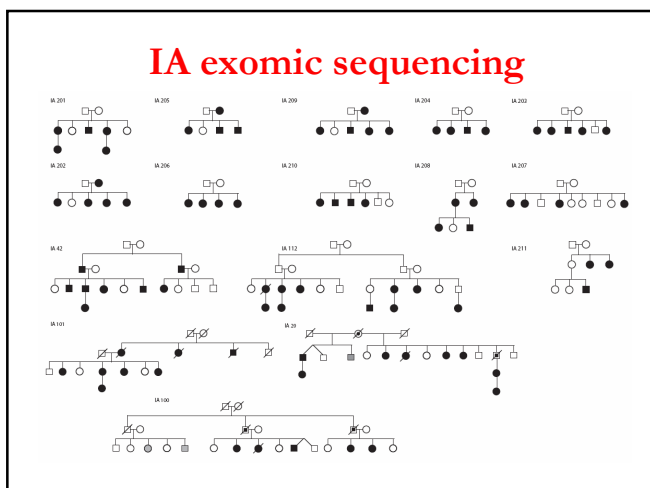
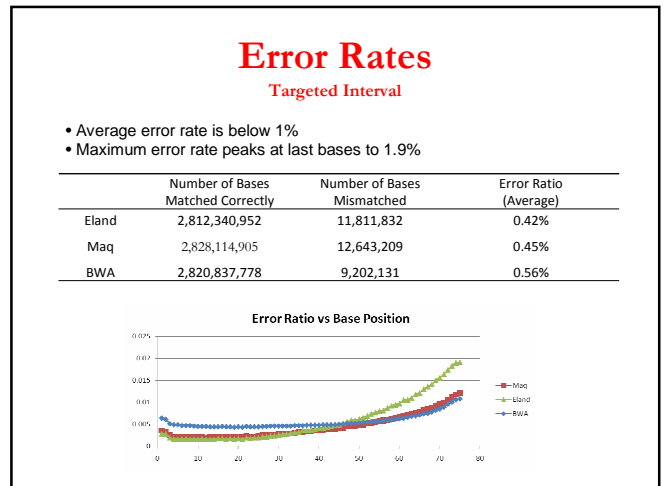
Targeted Interval

Over 95% of a total of 78,986,666 reads were mapped to human genome. Approximately 44% were within the target region.

|       | % Alignment to genome | % Alignment to TR | Coverage Depth |
|-------|-----------------------|-------------------|----------------|
| Eland | 98.68%                | 45.43% (44.83%)   | 92.77X         |
| Maq   | 94.23%                | 47.55% (44.80%)   | 92.72X         |
| BWA   | 98.68%                | 45.30% (44.70%)   | 92.50X         |

Within the ~ 28.6 million base-pairs targeted, more than 95% of bases were sequenced at least 5 times

|       | Eland  | Maq    | BWA    |
|-------|--------|--------|--------|
| ≥ 5X  | 95.45% | 95.50% | 95.55% |
| ≥ 10X | 91.21% | 91.30% | 91.37% |
| ≥ 20X | 83.52% | 83.61% | 83.69% |



- ## IA Formation/Rupture Risk
- **Preclinical diagnosis**
    - common + (rare?) + very rare variants
      - common variants: 4 to 7 fold increase in risk (explains ~10% of genetic risk)
      - contribution of rare variants
      - GWAS-4: sub-phenotype analysis
    - **Genetic + Environmental + Hemodynamic Factors**
    - smoking + hypertension
    - arterial flow patterns
      - branching, asymmetry
  - **New Therapeutic Approaches**