

Encode:

Encyclopedia of DNA elements in the human genome

over 400 scientists — over 200 million dollars

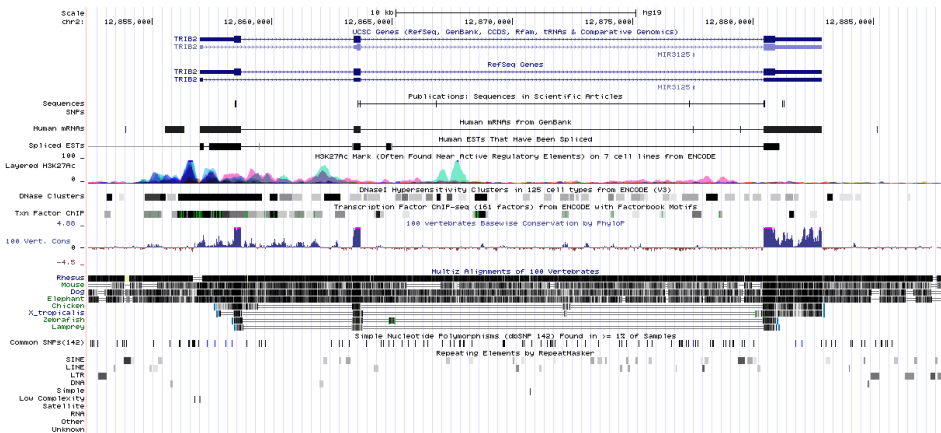
over 50 publications in 2012 and since

a huge amount of data

but the interpretation ?

Fanyan Pauli⁴, Kate R. Rosenbloom¹⁵, Peter Sabo¹⁶, Alexis Safi¹⁷, Amartya Sanyal¹, Noam Shores¹⁸, Jeremy M. Simon¹⁹, Linyun Song²⁰, Nathan D. Trinklin²¹, **Lead analysts (data production and analysis)** Edwin Birney²², James Brown²³, Chao Charo²⁴, Sarah DeBali²⁵, Kijung Dong²⁶, Ian Dunham²⁷, Jason Ernst²⁸, Terence S. Furey², Mark Gerstein³, Belinda Gardine³⁹, Melissa Green⁴⁰, Ross C. Hardson^{25,26}, Robert S. Harris²⁵, Javier Herrero⁴, Michael M. Hoffman¹⁶, Sowmya Iyer²⁷, Manolis Kellis¹⁹, Jaibab Khatun¹⁹, Poyaa Kheradpour¹⁹, Anshul Kundaje¹, Timo Lassmann²⁹, Qunhua Li²⁰, Xinying Luo¹, Georgi K. Marinov²⁹, Angelika Merkel²², Ali Morazzani³⁰, Stephen C.J. Parker³¹, Timothy E. Reddy³², Joel Rozowsky³³, Felix Schlesinger³⁴, Robert E. Thurman³⁵, Jie Wang³⁶, Lucas D. Ward³⁷, Troy W. Whitfield³⁸, Steven P. Wilder³⁹, Michael F. Young⁴⁰, Yip Hai Zhuang⁴¹, **Writing group** Bradley E. Bernstein³³, Edwin Birney²², Ian Dunham²⁷, Eric D. Green⁴², Chris Gunter⁴³, Michael Snyder¹², **NHGR project management (scientific management)** Michael J. Pazin¹⁹, Rebecca F. Lowdon³⁴, Laura A. L. Dillon³⁵, Leslie B. Adams³⁶, Caroline J. Kelly³⁷, Julia Zhang³⁸, Judith R. Wexler³⁹, Eric D. Green⁴², Peter J. Good⁴³, Elise A. Feingold⁴⁴, **Principal investigators (steering committee)** Bradley E. Bernstein³³, Edwin Birney²², Gregory E. Crawford¹⁷, Job Dekker¹, Laura Elnitski³⁴, Peggy J. Farnham³⁵, Mark Gerstein³, Morgan C. Giddings³⁸, Thomas R. Gingeras³⁹, Peic D. Green⁴⁰, Roderic Guigo⁴⁵, Ross C. Hardson^{25,26}, Timothy Hubbard⁴⁶, Manolis Kellis¹⁹, W. James Kent¹⁹, John A. D. Lieb¹⁹, Elliott H. Margulies¹⁵, Richard M. Myers¹⁴, Michael Snyder¹², John A. Stamatoyannopoulos⁴⁷, Scott A. Tenenbaum⁴⁸, Zhiping Weng¹⁹, Kevin P. White⁴¹, Barbara Wolf^{29,42}, **Boise State University and University of North Carolina at Chapel Hill Proteomics groups (data production and analysis)** Jaibab Khatun¹⁹, Yanbo Yu¹⁹, John Wrobel¹⁹, Brian A. Rsk¹⁹, Harsha P. Gunawardena⁴⁹, Heather C. Kuper⁵⁰, Christopher W. Maer⁵¹, Ling Xie⁵², Xian Chen⁵³, Morgan C. Giddings³⁸, **Broad Institute Group (data production and analysis)** Bradley E. Bernstein³³, Charles B. Epstein⁵⁴, Jason Ernst²⁸, Poyaa Kheradpour¹⁹, Tarjei S. Mikkelsen⁵⁵, Shawn Gillespie⁵⁶, Alon Corcos⁵⁷, Oren Ram⁵⁸, Xiaolan Zhang¹, Li Wang⁵⁹, Robbyn Issner⁶, Michael J. Coyne⁶⁰, Timothy Durham⁶¹, Manching Ku⁶², Thanh Truong⁶³, Lucas D. Ward³⁷, Robert C. Atshuler¹⁹, Matthew L. Eaton¹⁹, Manolis Kellis¹⁹, **Cold Spring Harbor, University of Geneva, Center for Genomic Regulation, Barcelona, RIKEN, Sanger Institute, University of Lausanne, Genomes Institute of Singapore group (data production and analysis)** Sarah DeBali²⁵, Carrie A. Davis⁴, Angelika Merkel²², Alec Dobin¹, Timo Lassmann²⁹, Ali Morazzani³⁰, Andrea Tanzer²², Julien Lagarde²², Wei Liu¹, Felix Schlesinger³⁴, Chenghai Xue⁶⁴, Georgi K. Marinov²⁹, Jaibab Khatun¹⁹, Brian A. Williams²⁹, Chris Zaleski¹, Joel Rozowsky³³, Maik Rödel²², Felix Kokocinski⁶⁵, Rehab F. Abdelhamid²⁸, Tyler Alkoti^{28,44}, Igor Antokhechkin¹, Michael T. Baer¹, Philippe Batut¹, Ian Bell⁶⁶, Kimberly Bell¹, Sudipto Chakraborty¹, Xian Chen¹, Jacqueline Christensen¹, Joao Curado⁶⁷, Thomas Derrien⁶⁸, Igor Drenkow¹, Erica Dumais⁶⁹, Jackie Durmas⁷⁰, Radha Duttagupta⁷¹, Megan Fastuca¹, Kota Fees-Toth⁷¹, Pedro Ferreira⁷², Sylvain Foissac⁷³, Melissa J. Fullwood⁴⁷, Yu Gao⁴⁴, David Gaidatzis⁷⁴, Assaf Gordon⁷⁵, Harsha P. Gunawardena⁴⁹, Cedric Hould¹⁶, Sonali Jha⁷⁶, Roy Johnson²⁷, Philipp Kapranov⁴⁰, Brandon King⁷⁷, Colin Kingswood⁷⁸, Guoliang Li⁷⁹, Oscar J. Luo¹, Eddie Park⁸⁰, Jonathan B. Preall¹, Kimberly Presaud⁸¹, Paolo Ribeca^{24,44}, Brian A. Rsk¹⁹, Daniel Roby⁴⁹, Xiaolan Ruan⁴⁷, Michael Sammeth^{24,44}, Kuljeet Singh Sandhu⁸², Lorin Schaeffer⁸³, Lei-Hoon See⁸⁴, Atif Shahab⁸⁵, Jorgen Skanke²², Ana Maria Suzuki⁸⁶, Hazuki Takahashi⁸⁷, Hagen Tilgner¹, Diane Trout¹⁹, Nathalie Walters⁷⁹, Huaien Wang¹, John Wrobel¹⁹, Yanbo Yu¹⁹, Yoshhide Hayashizaki⁸⁸, Jennifer Harrow³, Mark Gerstein³, Timothy J. Hubbard⁴⁶, Alexander Raymond²⁰, Sylvain E. Antonarakis⁸⁹, Gregory J. Hannon⁹⁰, Morgan C. Giddings³⁸, Yijun Yuan⁹¹, Barbara Wolf^{29,42}, Piero Carninci²⁹, Roderic Guigo⁴⁵, Thomas R. Gingeras³⁹, **Data coordination center at UC Santa Cruz (production data coordination)** Kate R. Rosenbloom¹⁵, Cricket A. Sloan⁹², Katrina Learned¹⁵, Venkit S. Malladi¹⁵, Matthew C. Wong¹⁵, Gall P. Barber¹⁵, Melissa S. Cline¹⁵, Timothy R. Dreszer¹⁵, Steven G. Heitner¹⁵, Donna Karolchik¹², W. James Kent¹⁹, Vanessa M. Kirkup¹⁵, Laurence R. Meyer¹⁵, Jeffrey C. Long¹⁵, Morgan Madden¹⁵, Brian J. Raney¹⁵, **Duke University, EBI, University of Texas, Austin, University of North Carolina-Chapel Hill group (data production and analysis)** Terence S. Furey², Linyun Song²⁰, Linda L. Grassefield⁹³, Paul G. Sires¹⁸, Bum-Joo Lee⁹⁴, Ann Battchenhouse¹, Nathan C. Sheffield¹, Jeremy M. Simon¹⁹, Kimberly A. Showers¹⁸, Alexis Safi¹⁷, Darin London¹⁷, Ashay A. Bhinge¹³, Christopher Shestak¹⁸, Matthew R. Schaner¹⁸, Seul Ki Kim¹⁸, Zhuzhu Z. Zhang¹⁸, Piotr A. Mieczkowski⁹⁵, Joanna O. Maciejowska¹⁸, Zheng Liu¹⁸, Ryan M. McDaniels¹⁸, Yunyun Ni¹⁸, Nam U. Rashid¹⁸, Jin Ja Kim¹⁸, Shera Adang¹⁸, Zhancheng Zhang¹⁸, Tianyuan Yang¹⁸, Deborah Writer¹⁸, Damian Kiedel¹⁸, Eason D. Lytle¹⁸, Vishweshwar R. Iyer¹⁸, Gregory E. Crawford¹⁷, **Genomes Institute of Singapore group (data production and analysis)** Guoliang Li⁷⁹, Kuljeet Singh Sandhu⁸², Weizhen Zheng⁸², Ping Wang⁸², Oscar J. Luo¹, Atif Shahab⁸⁵, Melissa J. Fullwood⁴⁷, Xiaolan Ruan⁴⁷, Yijun Yuan⁹¹, **AlphaGenome Institute, CNR-IRP-INA, Stanford group (data production and analysis)** Duhaob

Frankish⁹⁶, James Gilbert⁹⁷, Jose Manuel Gonzalez⁹⁸, Ed Griffiths⁹⁸, Rachel Harte¹⁵, David A. Hendrix⁹⁹, Cedric Hould¹⁶, Toby Hunt¹⁰⁰, Irwin Jungreis¹⁰¹, Mike Kay¹⁰², Edta Kasper¹⁰³, Michael F. Lin¹⁰⁴, Jian Liu¹⁰⁵, James Lovell¹⁰⁶, Jun Luo¹⁰⁷, Deepa Manthrapad¹⁰⁸, Marco Mariani¹⁰², Jonathan Mudge¹⁰⁹, Gaurab Mukherjee¹⁰², Cedric Notredame²², Baokang Pei²¹, Jose Manuel Rodriguez²⁵, Gary Saunders¹¹⁰, Andrea Sooner²⁰, Stephen Searle¹¹¹, Cristina Sisu²¹, Catherine Song¹¹², Charlie Steward¹¹³, Andrea Tanzer²², Electra Tapanari¹¹⁴, Michael L. Treiss¹¹⁵, Marijke J. van Baren¹¹⁶, Nathalie Walters⁷⁹, Stefan Washietl¹¹⁷, Laurens Wilmink¹¹⁸, Ammonia Zaidonis¹¹⁹, Zhongdong Zhang²⁰, Michael Brent²⁰, David Haussler¹²⁰, Manolis Kellis¹⁹, Alfonso Valencia¹²¹, Mark Gerstein³, Alexander Raymond²⁰, Roderic Guigo⁴⁵, Jennifer Harrow³, Timothy J. Hubbard⁴⁶, **Yale University, Harvard, University of Massachusetts Medical School, University of Southern California/UC Davis group (data production and analysis)** Stephen G. Landt¹²², Seth Frieze¹²³, Alexei Abzov¹²⁴, Nick Adelman¹²⁵, Roger P. Alexander¹²⁶, Raymond K. Auerbach¹²⁷, Suganthi Balasubramanian¹²⁷, Keith Bettinger¹²⁸, Ntin Bhardwaj¹²¹, Alan P. Boyle¹²⁹, Aina R. Cao⁶², Philip Caying¹²², Alexandra Charos⁶³, Yong Cheng¹²², Chao Cheng¹²², Catherine Eastman¹²¹, Ghia Euskirchen¹²², Joseph D. Fleming¹²², Fabian Grubert¹²², Lukas Habegger¹²¹, Manoj Hariharan¹²¹, Arif Harmanzi¹²¹, Sushma Iyengar¹²¹, Victor X. Jin¹²¹, Konrad J. Karczewski¹²², Hugo Lamb¹²², Phil Lachy¹²², Margy Lam¹²², Nathan Lammare-Vincant¹²², Jing Leng¹²¹, Jian Lin⁶⁸, Marianne Lindahl-Alier¹²⁴, Renjing Min¹²⁴, Benoit Miotto¹²⁴, Hannah Moranhan⁶², Zariq Motaqader¹²⁴, Xinming J. Mu²¹, Henriette O'Geen⁶², Zhengqing Ouyang¹²², Dorrelynn Patacila¹²², Baokang Pei²¹, Debasish Raha⁶³, Lucia Ramirez¹²³, Brian Reed¹²³, Joel Rozowsky³³, Andrea Sboner⁵⁸, Minkyi Shu¹²², Cristina Sisu²¹, Teri Sifler¹²², Heather Wet¹²², Linfeng Wu¹²², Xiaojin Xu¹²², Koon-Kiu Yan²¹, Xinqiong Yang¹²², Kevin Y. Yip²¹, Zhongdong Zhang²⁰, Kiem Struhl⁶⁴, Sherman M. Weissman¹, Mark Gerstein³, Peggy J. Farnham³⁵, Michael Snyder¹², **University of Alabama at Huntsville SUNT group (data production and analysis)** Scott A. Tenenbaum⁴⁸, **University of Pennsylvania, University of Chicago, Stanford group (data production and analysis)** Subhadip Karmakar¹²¹, Stephen G. Landt¹²², Raj R. Bhanadja¹²¹, Ailina Choudhury¹²¹, Marc Domanus¹²¹, Lijia Ma¹²¹, Jennifer Moran¹²², Dorrelynn Patacila¹²², Teri Sifler¹²², Alec Victorson¹²¹, Xinqiong Yang¹²², Michael Snyder¹², Kevin P. White⁴¹, **University of Heidelberg group (targeted experimental validation)** Thomas Auer⁶⁹, Lazaro Centanin⁶⁹, Michael Eichenlaub⁶⁹, Franziska Gruh⁶⁹, Stephan Heermann⁶⁹, Burkhard Hoekendorf⁶⁹, Daigo Inoue⁶⁹, Tanja Kellner⁶⁹, Stephan Kirchmair⁶⁹, Claudia Mueller⁶⁹, Robert Reinhardt⁶⁹, Lea Scherf⁶⁹, Stephanie Schneider⁶⁹, Rebecca Sinn⁶⁹, Beate Wittbrodt⁶⁹, **University of Massachusetts Medical School Bioinformatics group (data production and analysis)** Zhiping Weng¹⁹, Troy W. Whitfield³⁸, Jie Wang³⁶, Patrick J. Collins¹², Shelley F. Aldred¹², Nathan D. Trinklin²¹, Eric Christopher Partridge¹², Richard M. Myers¹⁴, **University of Massachusetts Medical School Genome Forming group (data production and analysis)** Job Dekker¹, Gaurav Jain¹, Bryan R. Lajoie¹, Amartya Sanyal¹, **University of Washington, University of Massachusetts Medical Center group (data production and analysis)** Gayathri Balasundaram¹³⁰, Daniel L. Bates¹³¹, Roshni Bhat¹³², Thea K. Canfield¹³², Morgan A. DeGisi¹³², Douglas Dunn¹³², Abigail K. Ebersole¹³², Tristan Frum¹³², Kavita Gang¹³², Erica Galt¹³², R. Scott Hansen¹³², Lisa Boatman¹³¹, Eric Haugen¹³¹, Richard Humbert¹³², Gaurav Jain¹, Andrew K. Johnson¹³², Ericka M. Johnson¹³², Tatyana V. Kutyavina¹³², Bryan R. Lajoie¹³², Kristen Lee¹³², Dimitra Lotakis¹³², Matthew T. Maunro¹³², Shane J. Neple¹³², Fiedenco V. Neri¹³², Eric D. Nguyen¹³², Hongzhu Qu¹⁶, Alex P. Reynolds¹³², Vaughn Rauch¹³², Eric Rynes¹³², Peter Sabo¹⁶, Minerva E. Sanchez¹², Richard S. Sandstrom¹³¹, Amartya Sanyal¹, Anthony O. Shafiq¹³², Andrew B. Stergachis¹³², Sean Thomas¹³², Robert E. Thurman³⁵, Benjamin Vernot¹³², Jeff Vong¹³², Thea K. Canfield¹³², Has Wang¹³², Yongqi Yang¹³², Miaochun Zhang¹³², Joshua M. Akey¹³², Michael Bender¹³², Michael O. Dorschner¹³², Mark Groutine¹³², Michael J. MacCoss¹³², Patrick Nawas¹³², George Stamatoyannopoulos¹³², Rajinder Kauf¹³², Job Dekker¹, John A. Stamatoyannopoulos⁴⁷, **Data Analysis Center (data analysis)** Ian Dunham²⁷, Kathryn Beal¹, Avish Brizma¹, Paul Flicek¹, Javier Herrero⁴, Nathan Johnson¹, Damian Keefe¹, Margus Luik¹, Nicholas M. Luscombe⁷⁹, Daniel Sobral¹, Juan M. Vazquez⁷⁹, Steven P. Wilder³⁹, Serafim Batzoglou¹²³, Arend Sidow¹²³, Nadine Hussam¹²³, Sofia Kyriazopoulou-Fanagiotopoulou¹²³, Max W. Libbrecht¹²³, Marco Schaub¹²³, Anshul Kundaje¹, Ross C. Hardson^{25,26}, Webb Miller²⁵, Belinda Gardine³⁹, Robert S. Harris²⁵, Weiseng Wang²⁵, Peter Bickel¹⁰⁰, Balazs Bantus¹⁰⁰, Nathan P. Boley¹⁰⁰, James B. Brown¹⁰⁰, Haiyan Huang¹⁰⁰, Qunhua Li²⁰, Jingyi Jin¹⁰⁰, William Stafford Noble¹⁰⁷, Jeffrey A. Billes¹⁰⁸, Orton J. Buske¹⁰⁸, Michael M. Hoffman¹⁶, Avinash D. Sahu¹⁰⁷, Peter V. Kharchenko¹⁰⁷, Peter J. Park¹, Dannon Baker¹⁰⁹, James Taylor¹⁰⁹, Zhiping Weng¹⁹, Sowmya Iyer²⁷, Xianjun Dong¹⁰⁹, Melissa Green⁴⁰, Xinying Luo¹, Jie Wang³⁶, Hualin Si¹², Jiali Zhuang³, Mark Gerstein³, Roger P. Alexander¹²⁶, Suganthi Balasubramanian¹²⁷, Chao Cheng¹²², Arif Harmanzi¹²¹, Lucie Lochovsky¹²¹, Renjing Min¹²⁴, Xinming J. Mu²¹, Joel Rozowsky³³, Koon-Kiu



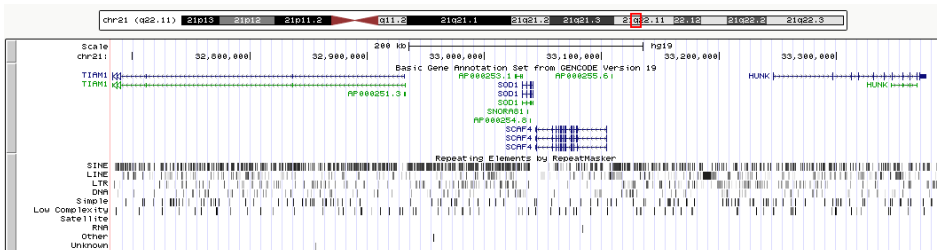
Expression refresh

- | | | | | | |
|--|--|--|---|---|--|
| Affy Exon Array
hide | Affy GNF1H
hide | <input checked="" type="checkbox"/> Affy RNA Loc
hide | Affy U133
hide | Affy U133Plus2
hide | Affy U95
hide |
| Allen Brain
hide | Burge RNA-seq
hide | <input checked="" type="checkbox"/> CSHL Small RNA-seq
hide | <input checked="" type="checkbox"/> ENC Exon Array...
hide | <input checked="" type="checkbox"/> ENC ProtGeno...
hide | <input checked="" type="checkbox"/> ENC RNA-seq...
hide |
| <input checked="" type="checkbox"/> GIS RNA PET
hide | GNF Atlas 2
hide | GWIPS-viz Riboseq
hide | <input checked="" type="checkbox"/> Illumina WG-6
hide | PeptideAtlas
hide | qPCR Primers
hide |
| <input checked="" type="checkbox"/> RIKEN CAGE Loc
hide | <input checked="" type="checkbox"/> Sestan Brain
hide | | | | |

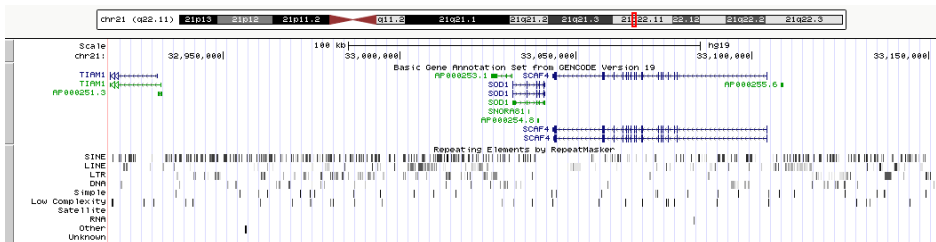
Regulation refresh

- | | | | | | |
|--|--|---|--|---|--|
| <input checked="" type="checkbox"/> ENCODE Regulation...
show | <input checked="" type="checkbox"/> CD34 Dnase1
hide | CpG Islands...
hide | <input checked="" type="checkbox"/> ENC Chromatin...
hide | <input checked="" type="checkbox"/> ENC DNA Methyl...
hide | <input checked="" type="checkbox"/> ENC DNase/FAIRE...
hide |
| <input checked="" type="checkbox"/> ENC Histone...
hide | <input checked="" type="checkbox"/> ENC RNA Binding...
hide | <input checked="" type="checkbox"/> ENC TF Binding...
hide | <input checked="" type="checkbox"/> FSU Repli-chip
hide | <input checked="" type="checkbox"/> Genome Segments
hide | <input checked="" type="checkbox"/> NKI Nuc Lamina...
hide |
| <input checked="" type="checkbox"/> ORegAnno
hide | <input checked="" type="checkbox"/> Stanf Nucleosome
hide | <input checked="" type="checkbox"/> SUNY SwitchGear
hide | <input checked="" type="checkbox"/> SwitchGear TSS
hide | TFBS Conserved
hide | TS miRNA sites
hide |
| UCSF Brain Methy
hide | UMMS Brain Hist
hide | <input checked="" type="checkbox"/> UW Repli-seq
hide | Vista Enhancers
hide | | |

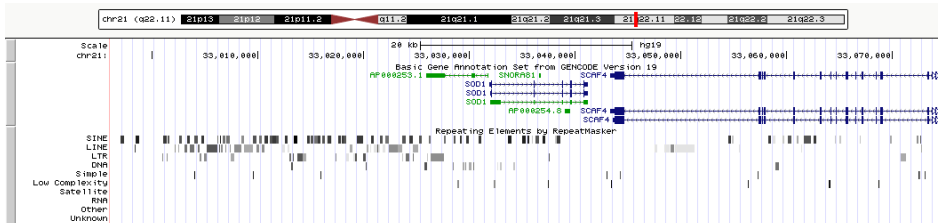
Repetitive Elements



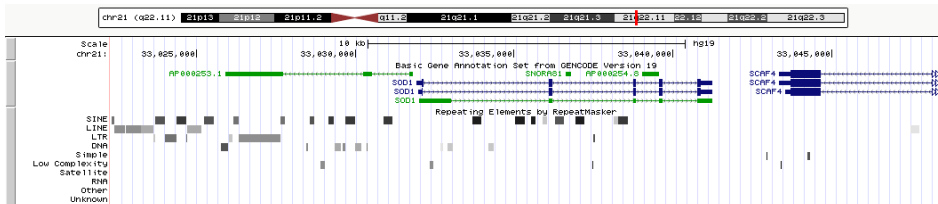
Repetitive Elements



Repetitive Elements



Repetitive Elements

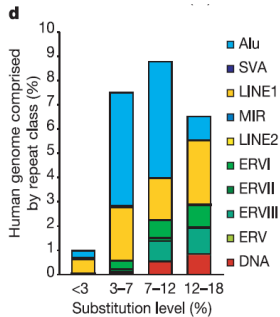
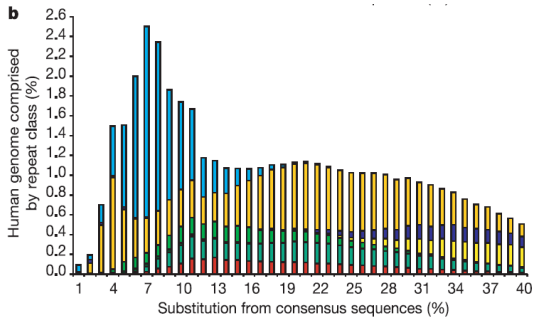


TE's - SINEs

Alu's are ~285bp long. They were anciently derived from an 7SL-RNA gene in some common ancestor of primates. They contain no protein coding regions but do have an internal polIII promoter (as does 7SL-RNA).

	Copies	bp	Genome
Alu	1,090,000	290,100,000	10.60%

TE's - History of Human element activity



TE's - SINEs

	Copies	bp	Genome
Alu	1,090,000	290,100,000	10.60%
MIR	393,000	60,100,000	2.20%
MIR3	75,000	9,300,000	0.34%
SINEs	1,558,000	359,600,000	13.14%