

Theoretical analysis indicates Genome is not a blueprint and oocytes have the instruction.

Koichi Itoh¹

1: The Institute for Theoretical Molecular Biology
21-13, Rokurokuso-cho, Ashiya, Hyogo, JAPAN, 659-0016
TEL: +81-797-35-6368
FAX: +81-797-35-6368
e-mail: koichiitoh@yahoo.co.jp

Corresponding author: Koichi Itoh
1: The Institute for Theoretical Molecular Biology
Room204, 2-9 Shinnouzuka-cho, Ashiya, Hyogo, JAPAN, 659-0016
TEL: +81-797-35-6368
FAX: +81-797-35-6368
e-mail: koichiitoh@yahoo.co.jp

Summary

Genome has been thought to be a blueprint, but what type of the blueprint has been a mystery. We have been thought that Genome is not a blueprint. Genome is storage of genes. An assembly instruction exists in oocytes. After fertilization, zygotes express genes according to the instruction. We get this result from the survey of Databases. Oocytes express several types of genes for blastula development and tissue differentiation such as Zinc finger proteins, Homeobox proteins, Pou homeobox proteins, Forkhead box proteins, High-mobility group proteins, Sex determination proteins, WNT and PAX. This result indicates that oocytes have the instruction. Finally, we foresee that once organogenesis begins tissue differentiation proceeds autonomously and human bodies are built.

After Human Genome Project was over and human genome sequence was determined, we expected that a blueprint existed in the genome. However we could not find any traces of the blueprint. 1) If Genome is a blueprint, it must have the regularity in the sequence. But there is no regularity. For example, within Genome, there are a lot of pseudogenes. 2) In glycolysis pathway enzymes, hexokinase1 is located in 10q22, glucose-6-phosphate isomerase is located in 19q13.1, phosphofructokinase, liver type is located in 21q22.3, phosphofructokinase, muscle type is located in 12q13.3, phosphofructokinase, platelet type is located in 10p15.3-p15.2, aldolase is located in 16p11.2, triosephosphate isomerase is located in 12p13, glyceraldehydes-3-phosphate dehydrogenase is located in 12p13.31-p13.1, phosphoglycerate kinase is located in 6p21.1-p12, phosphoglycerate mutase is located in 10q25.3, enolase 1 is located in 1pter-p36.13, enolase 2 is located in 12p13, enolase 3 is located in 17pter-p12, pyruvate kinase, liver and red blood cell type is located in 1q21 and pyruvate kinase muscle type is located in 15q22. You can find easily a lot of examples like this in other pathways. In glycolysis, important genes for enzymes are not located in the near portion on Genome. These 2 examples are enough for show you that Genome is not a blueprint. We have been thinking that the instruction exists in oocytes. Unfortunately, at that time, methods for massive gene expression profiling and Databases were premature and we could not get enough data to answer the question. But now owing to the spread of microarray analysis and large scale expression profilings, we can easily get the data we want and Databases pile up enormous data.

Database survey

In 2006, Kocabas *et al.* reported the gene expression profile of human oocytes¹. In that profile, we paid attention for genes for transcription. We thought that among those genes, genes for instruction existed. We surveyed those genes in NCBI (National Center for Biotechnology) Databases. We found genes for 1) Zinc finger proteins are large and diverse group of proteins in which the DNA-binding region contains projections (“fingers”) with Cys and/or His residues folding around zinc atom. Some examples are GATA factors important in the development of the blood and the gut, Kruppel in the early *Drosophila* embryo, Krrox20 in the rhombomeres of the hindbrain. 2) Pou homeobox proteins are related to pituitary development and neural fate. 3) Homeobox proteins have a special role in the control of anteroposterior pattern in animals. 4) High-mobility group and Sex determination are related sex determination. 5) Forkhead box are related to head development and liver formation. 6) PAX proteins are related neural specification and eye

development²⁻⁷. LHX5 is related to control of differentiation and development of the forebrain. TBX3 plays a role in the anterior/posterior axis. TBX5 plays a role for heart development and specification of limb identity. TCF15 plays a role for transcriptional regulation of patterning of the mesoderm (Table1). These genes are essential for blastula development and tissue differentiation. And genes for DNA maintenance and remodeling, cell cycle, transcription factors or transcription factor subunits are also essential for the expression of genes for metabolism, cell growth, and cytoskeleton. You can find genes for Oct3/4, Sox2, Klf4 and c-Myc in <http://www.crl.msu.edu/Supp%20WEB/Kocabas%20et%20al%20Supp.%20Web%20Index.htm>. These four are essential for the generation of iPS cells⁸. We can indicate that the number of genes related in development is very large.

Discussion

Here we indicate that after fertilization, oocytes express genes for blastula development and tissue differentiation. We show the evidence that Genome is not the blueprint and oocytes have the instruction from the theoretical point of view. Kocabas *et.al* reported only up-regulated genes, that is why all essential genes for blastula development and tissue differentiation were not included in their lists. However, in the near future, experimental biologists will find all essential genes for blastula development and tissue differentiation. We surveyed Pathways for development in KEGG (Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>) and Transpath: The Pathway Database; <http://www.transpath.de/>), but we could not find any pathways for development. From now on, pathways for blastula development and tissue differentiation will be elucidated. Finally, we foresee that once organogenesis begins tissue differentiation proceeds autonomously and human bodies are built.

Reference

1. Kocabas, A. M. *et al.* (2006). The transcriptome of human oocytes. *Proc. Natl Acad. Sci. USA* 103, 14027-14032
2. Slack, J.M.W. *Essential Developmental Biology*, second edition, (2006) (Blackwell Publishing)
3. Gilbert, S.F. *Developmental Biology*, eighth edition, (2006) (Sinauer Association Inc.)
4. Alberts, B. *et al.* (2008). *Molecular Biology of the Cell*, Fifth edition (Garland Science)
5. Schoenwolf, G.C. *et al.* (2009). *Larsen's Human Embryology*, Fourth edition, (Churchill Livingstone)
6. Wolpert, L. *et al.* (2007). *Principles of Development*, Third edition, (Oxford University Press)
7. Moody, S.A. (2007) *Principles of Developmental Genetics* (Academic Press)
8. Takahashi, K. *et al.* (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861-72

Table1. Up-regulated genes in human Oocytes

Unigene	Symbol	Gene name	Group
Hs.37706	ZCCHC8	Zinc finger, CCHC domain containing 8	Zinc finger
Hs.523710	ZDHHC18	Zinc finger, DHHC-type containing 18	
Hs.27239	ZDHHC5	Zinc finger, DHHC-type containing 5	
Hs.592065	ZDHHC7	Zinc finger, DHHC-type containing 7	
Hs.612084	ZHX1	Zinc fingers and homeoboxes 1	
Hs.507355	ZNF10	Zinc finger protein 10	
Hs.479874	ZNF136	Zinc finger protein 136	
Hs.592591	ZNF148	Zinc finger protein 148	
N.D.	ZNF161	N.D.	
Hs.155204	ZNF174	Zinc finger protein 174	
Hs.172979	ZNF177	Zinc finger protein 177	
Hs.112556	ZNF202	Zinc finger protein 202	
Hs.530988	ZNF237	zinc finger protein 237 isoform 1	
Hs.59757	ZNF281	Zinc finger protein 281	
Hs.484324	ZNF354A	Zinc finger protein 354A	
Hs.659797	ZNF558	Zinc finger protein 558	
Hs.655107	ZNF559	Zinc finger protein 559	
Hs.13323	ZNF574	Zinc finger protein 574	
Hs.709730	ZNF586	Zinc finger protein 586	
Hs.388024	ZNF76	Zinc finger protein 76 (expressed in testis)	
Hs.38004	ZNF77	Zinc finger protein 77	
Hs.37138	ZNF85	Zinc finger protein 85	
Hs.388927	YY1	YY1 transcription factor	
Hs.518249	ZNF9	CCHC-type zinc finger, nucleic acid binding protein	
Hs.493649	POU2F1	POU class 2 homeobox 1	Pou homeobox
Hs.182505	POU3F2	POU class 3 homeobox 2	
Hs.654522	POU4F1	POU class 4 homeobox 1	
Hs.249184	POU5F1	POU class 5 homeobox 1	
Hs.67397	HOXA1	Homeobox A1	Homeobox
Hs.592172	HOXA13	Homeobox A13	

Hs.660918	HOXA7	Homeobox A7		
Hs.83465	HOXD1	Homeobox D1		
Hs.152414	HOXD13	Homeobox D13		
Hs.714377	HMG20A	High-mobility group 20A	High-mobility group	
Hs.406534	HMG20B	High-mobility group 20B		
Hs.588815	HMG2L1	High-mobility group protein 2-like 1		
Hs.518805	HMGA1	High mobility group AT-hook 1		
Hs.434953	HMGB2	High-mobility group box 2		
Hs.555947	LEF1	Lymphoid enhancer-binding factor 1		
Hs.236774	HMGN4	High mobility group nucleosomal binding domain 4		
Hs.201671	SOX13	SRY (sex determining region Y)-box 13		Sex determination
Hs.95582	SOX15	SRY (sex determining region Y)-box 15		
Hs.529462	SOX30	SRY (sex determining region Y)-box 30		
Hs.657542	SOX5	SRY (sex determining region Y)-box 5		
Hs.546573	FOXD3	Forkhead box D3	Forkhead box	
Hs.289292	FOXL2	Forkhead box L2		
Hs.239	FOXM1	Forkhead box M1		
Hs.370666	FOXO1A	Forkhead box O1		
	FOX3A	Forkhead box 3A		
Hs.29764	WNT6	wingless-type, member 6	WNT	
Hs.443881	PAXIP1	PAX interacting protein 1	paired box (PAX) gene	
Hs.654389	CUTL1	Cut-like homeobox 1	homeodomain family of DNA binding protein	
Hs.162032	HBP1	HMG-box transcription factor 1	HMG-box transcription factor 1	
Hs.302029	LHX5	LIM homeobox 5	control of differentiation and development of the forebrain	
Hs.129895	TBX3	T-box 3	play a role in the anterior/posterior axis	
Hs.381715	TBX5	T-box 5	heart development and specification of limb identity.	
Hs.437	TCF15	Transcription factor 15 (basic helix-loop-helix)	transcriptional regulation of patterning of the mesoderm	
Hs.33102	TFAP2B	Transcription factor AP-2 beta	developmentally regulated activator	
Hs.696032	PPARD	Peroxisome proliferator-activated receptor delta	nuclear hormone receptor	
Hs.655020	MXD4	MAX dimerization protein 4	regulation of cell growth in differentiating tissues	
Hs.517948	DHX30	DEAH (Asp-Glu-Ala-His) box polypeptide 30	putative RNA helicases	
Hs.191518	DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9		

Hs.397465	HIPK2	homeodomain interacting protein kinase 2	serine/threonine nuclear kinase
Hs.471991	MTF1	Metal-regulatory transcription factor 1	Metal-regulatory transcription factor
Hs.483616	RNF14	Ring finger protein 14	ubiquitination
Hs.561815	STAU2	Staufen, RNA binding protein, homolog 2	RNA binding protein
Hs.285354	MAX	MYC associated factor X	cell proliferation, differentiation and apoptosis
Hs.638121	CRK	V-crk sarcoma virus CT10 oncogene homolog	signal transduction
Hs.437075	CREB5	CAMP responsive element binding protein 5	CRE (cAMP response element)-binding protein
Hs.654952	POLR2J	Polymerase (RNA) II polypeptide J, 13.3kDa	RNA polymerase II subunit
Hs.567494	YBX2	Y box binding protein 2	Y box family of nucleic acid-binding proteins
Hs.385998	WDHD1	WD repeat and HMG-box DNA binding protein 1	signal transduction/DNA maintenance and remodeling
Hs.26516	ASF1B	ASF1 anti-silencing function 1 homolog B	DNA maintenance and remodeling
Hs.162233	CHD4	Chromodomain helicase DNA binding protein 4	
Hs.202672	DNMT1	DNA (cytosine-5-)-methyltransferase 1	
Hs.713611	DNMT3B	DNA (cytosine-5-)-methyltransferase 3 beta	
Hs.368322	CDH8	chromodomain helicase DNA binding protein 8	
Hs.144287	HEY2	hairy/enhancer-of-split related with YRPW motif 2	
Hs.368971	NCOA6	Nuclear receptor coactivator 6	
Hs.589489	SMARCA5	SWI/SNF related, matrix associated, member 5	
Hs.558463	MSX2	Spen homolog, transcriptional regulator (Drosophila)	
Hs.184298	CDK7	Cyclin-dependent kinase 7	Cell cycle
Hs.436975	CLOCK	circadian locomotor output cycles kaput protein	
Hs.79353	TFDP1	Transcription factor Dp-1	
Hs.348418	DR1	down-regulator of transcription 1	transcription factor OR transcription factor subunit
Hs.654393	E2F1	E2F transcription factor 1	
Hs.591015	ELK3	ELK3, ETS-domain protein (SRF accessory protein 2)	
Hs.25647	FOS	V-fos FBJ murine osteosarcoma viral oncogene homolog	
Hs.654350	GABPB2	GA binding protein transcription factor, beta subunit 2	
Hs.406157	HSF2BP	heat shock transcription factor 2 binding protein	
Hs.534074	NFATC1	Nuclear factor of activated T-cells,calcineurin-dependent 1	
Hs.632209	NFATC3	Nuclear factor of activated T-cells,calcineurin-dependent 3	
Hs.631886	REL	V-rel reticuloendotheliosis viral oncogene homolog	
Hs.514276	SP2	Sp2 transcription factor	

Hs.443258	SREBF2	Sterol regulatory element binding transcription factor 2	
Hs.153088	TAF1A	TATA box binding protein (TBP)-associated factor	
Hs.122752	TAF2	TATA box binding protein (TBP)-associated factor	
Hs.18857	TAF4	TATA box binding protein-associated factor	
Hs.590872	TBP	TATA box binding protein	
Hs.486507	TBPL1	TBP-like 1	
Hs.26837	TRIM33	Tripartite motif-containing 33	
Hs.460988	CBFB	Core-binding factor, beta subunit	
Hs.555985	NARG1	NMDA receptor regulated 1	
Hs.440219	UBN1	Ubiquitin 1	
Hs.573153	TCF7	Transcription factor 7 (T-cell specific, HMG-box)	
N.D.	TCFL1	N.D.	
Hs.12229	KLF11	Kruppel-like factor 11	
Hs.428027	PBX3	Pre B-cell leukemia homeobox 3	
Hs.92236	MLL4	Myeloid/lymphoid or mixed-lineage leukemia 4	
Hs.90753	HTATIP2	HIV-1 Tat interactive protein 2, 30kDa	
Hs.528641	SIRT7	Sirtuin (silent mating type information regulation 2 homolog) 7	unknown
Hs.271940	ELF4	E74-like factor 4 (ets domain transcription factor)	
Hs.463045	GCN5L2	GCN5 general control of amino-acid synthesis 5-like 2 (yeast)	
Hs.530539	NFRKB	Nuclear factor related to kappaB binding protein	
Hs.127950	BRD1	Bromodomain containing 1	
Hs.286145	MED21	mediator complex subunit 21	
N.D.	SNARCB1	N.D.	
N.D.	SMRCC2	N.D.	

N.D. not detected in NCBI database.