



# Mutations and binding sites of human transcription factors

Frederick Kinyua Kamanu<sup>†</sup>, Yulia A. Medvedeva<sup>†</sup>, Ulf Schaefer, Boris R. Jankovic, John A. C. Archer and Vladimir B. Bajic\*

Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

## Edited by:

William Muir, Purdue University, USA

## Reviewed by:

Dahlia Nielsen, North Carolina State University, USA

Yunlong Liu, Indiana University, USA

## \*Correspondence:

Vladimir B. Bajic, Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom Saudi Arabia.

e-mail: vladimir.bajic@kaust.edu.sa

<sup>†</sup>Frederick Kinyua Kamanu and Yulia A. Medvedeva have contributed equally to this work.

Mutations in any genome may lead to phenotype characteristics that determine ability of an individual to cope with adaptation to environmental challenges. In studies of human biology, among the most interesting ones are phenotype characteristics that determine responses to drug treatments, response to infections, or predisposition to specific inherited diseases. Most of the research in this field has been focused on the studies of mutation effects on the final gene products, peptides, and their alterations. Considerably less attention was given to the mutations that may affect regulatory mechanism(s) of gene expression, although these may also affect the phenotype characteristics. In this study we make a pilot analysis of mutations observed in the regulatory regions of 24,667 human RefSeq genes. Our study reveals that out of eight studied mutation types, “insertions” are the only one that in a statistically significant manner alters predicted transcription factor binding sites (TFBSs). We also find that 25 families of TFBSs have been altered by mutations in a statistically significant manner in the promoter regions we considered. Moreover, we find that the related transcription factors are, for example, prominent in processes related to intracellular signaling; cell fate; morphogenesis of organs and epithelium; development of urogenital system, epithelium, and tube; neuron fate commitment. Our study highlights the significance of studying mutations within the genes regulatory regions and opens way for further detailed investigations on this topic, particularly on the downstream affected pathways.

**Keywords:** SNP, insertion, deletion, mutation, transcription factor, transcription factor binding site, promoter region, bioinformatics

## INTRODUCTION

Mutations in any genome may lead to phenotype characteristics that determine ability of an individual to cope with environmental challenges (Kopp and Hermisson, 2009). In studies of human biology there is a continuous effort to identify mutations and this data can be freely accessed from public repositories such as HapMap (Belmont et al., 2005), dbSNP (Sayers et al., 2010), and the 1000 Genomes project (Altshuler et al., 2010). It is well known that phenotypic specificities determine how individuals react to drugs (Batist et al., 2011; Callahan and Abercrombie, 2011), to infections (González-Hernández et al., 2010), or how predisposed they are to inherited diseases such as cystic fibrosis (Gu et al., 2009; Antigny et al., 2011), Huntington's disease (Roze, 2011), galactosemia (Bennett, 2010), et cetera. Most of the research in this field has been focused on the studies of non-synonymous mutation effects on the final gene products, peptides, and their alterations (Kumar et al., 2009). Several databases have been developed in order to facilitate the study of these genetic variations and their consequences. Resources such as the Human Gene Mutation Database (HGMD; Stenson et al., 2009) and the Online Mendelian Inheritance in Man (OMIM) database (Amberger et al., 2009) collect mutations occurring across the entire human genome, whereas other repositories such as the *FLCN* Gene Database (Lim et al., 2010) are locus specific.

However, a considerably less attention has been given to the mutations that may affect regulatory mechanism of gene

expression, although these may also affect phenotype characteristics. Previous studies of the interactions of mutations and regulatory processes have associated mutations within the promoter region of certain key genes, with an increased susceptibility to different disorders such as Pancreatic cancer (Hamacher et al., 2009), type 2 Diabetes (Song et al., 2009), Myelodysplastic syndrome (Ma et al., 2010), and Idiopathic Pulmonary Arterial Hypertension (Yu et al., 2009). Polymorphisms in the intragenic or regulatory regions may influence transcription factor (TF) binding to DNA or may affect gene splicing (Heckmann et al., 2010; Kasowski et al., 2010). A number of resources has been compiled to provide more easy insights into effects of mutations to transcription and gene regulation, such as those related to protein coding genes (Conde et al., 2006; Kim et al., 2008), or for miRNA (Bao et al., 2007; Hariharan et al., 2009; Alexiou et al., 2010; Schmeier et al., 2011).

Here we performed a large-scale pilot study of mutations within the regulatory regions of 24,667 human RefSeq (Pruitt et al., 2009) genes. Our study reveals for the first time that out of eight studied mutation types, “insertions” is the only type that alters the predicted transcription factor binding sites (TFBSs) in a statistically significant manner. We also identified 25 families of TFBSs have been altered by mutations in a statistically significant manner in the promoter regions we considered. The related TFs are prominent in processes related to intracellular signaling, cell fate, epithelium, morphogenesis of organs and epithelium, development of urogenital system, epithelium and tube, neuron fate commitment,

et cetera. These observations highlight the significance of studying mutations with the genes regulatory regions and opens way for further detailed studies on this topic, particularly on the downstream affected pathways.

## RESULTS

We analyzed promoter regions of 24,667 human genes from RefSeq for the presence of TFBSs using Transfac Professional ver. 11.4 (Matys et al., 2006) and we predicted 1,077,742 TFBSs in these promoter regions. At the same time we found 343,024 mutations associated with the same promoter regions based on data from dbSNP. Of these, 122,023 TFBSs were altered by 104,514 mutations. Details of the mapped mutations and TFBSs by chromosome are provided in Tables S1 and S2 in Supplementary Material, respectively.

We analyzed the following eight mutations: “Single,” “Insertion,” “Deletion,” “In-del,” “Multiple Nucleotide Polymorphism” (MNP), “Mixed,” “Named,” and “Microsatellite.” “Single” represents single nucleotide variation; with all observed alleles are single nucleotides (can have 2, 3, or 4 alleles). “Microsatellite” corresponds to the situation when the observed allele from dbSNP is a variation in counts of short tandem repeats. “Named” represent polymorphisms in presence of complex structures, as transposons, e.g., (Alu)/-. “Mixed” corresponds to the cluster containing submissions from multiple classes. MNP represent situation where the alleles are all of the same length, and length >1. “Insertion” corresponds an insertion relative to the reference assembly. “Deletion” corresponds to a deletion relative to the reference assembly. “In-del” corresponds to situation when both insertions and deletions relative to reference genome were found in particular position. Our analysis suggests that only the insertion type of mutations alters TFBSs in a statistically significant manner (Table 1). On the other hand, the analysis of TFBSs altered by mutations indicates that 25 TFBS types are altered in a statistically significant manner. We associated TFs to these 25 TFBS types and found that they comprise: *HNF3 alpha*, *Pax-2*, *Pax-3*, *Pax-4*, *Pax-5*, *Pax-6*, *AIRE*, *PLZF*, *myogenin/NF-1*, *ZNF219*, *FOX* factors, *STAT1*, *CHX10*, *HNF3 beta*, *c-Maf*, *Tax/CREB*, *FOXPI*, *MyoD*, “*c-Ets-1 p54*,” *KROX*, *DEAF1*, *VDR*, *CAR*, *PXR*, *FAC1*, *PPARalpha:RXRalpha*, *PPARgamma:RXRalpha*, and *Spz1*. We also found the initiator element, “Muscle initiator sequences-19,” is altered in a statistically significant manner (Table 2). Details including TFs UniProt IDs (Jain et al., 2009) are provided in Table S3 in Supplementary Material. We used GeneMANIA (Warde-Farley et al., 2010) to analyze processes where the TFs which bind to the above mentioned 25 TFBS types potentially exert their effects and found (Table S4 in Supplementary Material) that in addition to activities usually associated with TF functioning, they are also prominent in processes related to intracellular signaling, cell fate, epithelium, morphogenesis of organs and epithelium, development of urogenital system, epithelium and tube, neuron fate commitment, response to nutrient levels, developmental processes, response to extracellular stimulus.

## DISCUSSION

The global analysis of mutations within TFBSs makes space for more detailed insights into potential effects that such mutations

can produce. For example, a mutation within a transcription initiation regulatory region can make one of the following downstream effects:

- the ability of altered TFBS to bind the same TFs remains (though possibly with different affinity)
- altered TFBS cannot bind the original TFs
- altered TFBS can bind new TFs as well as the original TFs
- altered TFBS can only bind new TFs
- altered TFBS cannot bind any TF
- segment of a regulatory region that previously did not bind any TF, may acquire ability to bind some TFs, thus effectively new TFBSs are introduced.

Studying such different scenarios in particular cases may provide insights into mutation effects (see Heckmann et al., 2010). Schmeier et al. (2011) have developed a database that provides a possibility to explore such potential effects of interactions of mutations with TFBSs in the promoter regions of miRNAs.

Due to degenerative sequence properties of TFBS, one can assume that many SNPs that overlap TFBSs will not affect the ability of TF to bind them. At the same time, larger mutations (such as insertions, deletions, MNP, etc.) are more likely to significantly affect affinity of TF to bind such modified TFBSs, and likely to destroy TFBSs. So, negative selection would reduce frequency of long polymorphisms within TFBSs and for this reason, such long polymorphisms are not likely to be significantly enriched within TFBSs. This is why our finding that “insertions” are statistically significant is an interesting one. Further research is needed to suggest the reason behind such an observation.

We observed that TFBSs of two TF families, PAX and FOX, are prominently associated with mutations. TFs from PAX family are associated with tissue specific gene expression and linked to development of specific tissues, including kidney and optic nerves (PAX-2; Lindoso et al., 2009); ear, eye, and facial development (PAX-3; Zhang et al., 2012); pancreatic islet beta cells (PAX-4; Collombat et al., 2009); b-cell differentiation, as well as neural and spermatogenesis development (PAX-5; Decker et al., 2009) and eyes and sensory organs, certain neural, and epidermal tissues development (PAX-6; Guo et al., 2010; Rowan et al., 2010). Forkhead box (FOX) family of TFs is implicated in processes of embryonic development, cell growth, proliferation, and cell differentiation (Hannenhalli and Kaestner, 2009).

While insertion type of mutations was the only one that appeared statistically significantly altering TFBSs, the other seven types could be considered to be results of uniform random changes of the genome.

The results obtained provide the foundation to investigate potentially affected pathways that are controlled by the TFs binding the most affected TFBSs and provide us links to important biological processes related to these TFs.

## CONCLUSION

Our analysis revealed that “insertion” is only statistically significant type of mutations in the predicted TFBSs in the regulatory regions of human genes that we explored. Also, we singled out 25 TFBS families that in our analysis appear statistically significantly

**Table 1 | Enrichment of mutation types found altering predicted TFBSs.**

Mutation type	Total number of mutations altering TFBSs	Total number of mutations not altering TFBSs	Uncorrected <i>P</i> -value	Bonferroni corrected <i>P</i> -value
Insertion	9374	20716	3.56E−003	2.85E−002
Single	86040	195689	2.57E−002	2.05E−001
Mixed	442	918	5.53E−002	4.42E−001
Named	218	482	3.63E−001	2.90E+000
MNP	168	370	3.67E−001	2.93E+000
In-del	2014	4616	5.69E−001	4.56E+000
Microsatellite	9	53	9.99E−001	7.99E+000
Deletion	6249	15666	1.00E+000	8.00E+000

**Table 2 | Statistically enriched altered TFBSs.**

Transcription factor	Total number of altered TFBSs	Total number of not altered TFBSs	Uncorrected <i>P</i> -value	Bonferroni corrected <i>P</i> -value
PLZF	2721	10848	0.00E+000	0.00E+000
Pax-6	4671	27237	0.00E+000	0.00E+000
PPARgamma:RXRalpha, PPARgamma	5079	27438	0.00E+000	0.00E+000
AIRE	537	2172	0.00E+000	0.00E+000
myogenin/NF-1	6469	26368	0.00E+000	0.00E+000
STAT1	621	3639	0.00E+000	0.00E+000
FAC1	1583	10354	0.00E+000	0.00E+000
Pax-4	3942	8182	0.00E+000	0.00E+000
Pax-2	1431	9394	0.00E+000	0.00E+000
Pax-3	1942	9761	0.00E+000	0.00E+000
ZNF219	146	394	0.00E+000	0.00E+000
FOX factors	475	2348	0.00E+000	0.00E+000
Muscle initiator sequences-19	917	5364	3.74E−010	6.54E−008
DEAF1	205	972	3.94E−010	6.89E−008
PPARalpha:RXRalpha	2622	15957	6.72E−010	1.18E−007
VDR, CAR, PXR	1317	7229	7.40E−010	1.29E−007
Spz1	2911	15960	7.67E−010	1.34E−007
VDR	2841	17551	9.56E−010	1.67E−007
Pax-5	10529	43269	5.65E−009	9.88E−007
HNF3alpha	598	3701	1.06E−007	1.86E−005
Tax/CREB	2044	14374	3.04E−006	5.33E−004
c-Ets-1 p54	2749	19716	7.93E−006	1.39E−003
Lhx3	64	265	1.20E−005	2.09E−003
Arnt	358	2205	2.25E−005	3.94E−003
FOXP1	16	34	8.03E−005	1.41E−002

altered by mutations. These open possibility to further explore individual effects of the altered TFBSs and their downstream and upstream regulation networks that can pave way for insights into pathways and diseases potentially affected.

## MATERIALS AND METHODS

### DATA

#### Promoters

We extracted promoters regions of 24,667 human genes from RefSeq (Pruitt et al., 2009). Promoters covered the region of [−1000, +500] relative to 5′ end of gene. Human genome version hg19 from UCSC Genome Browser database (Fujita et al., 2011) is used.

#### Mutations

We downloaded 3,3026,121 SNPs from the UCSC Genome Browser database (Fujita et al., 2011). These SNPs are derived from dbSNP build 132 (Sayers et al., 2010) and are available on the hg19 assembly of the human genome. The resulting set contains 319,820 polymorphisms. Based on genomic coordinates we identified SNPs that overlap promoter sequences, as well as those altering predicted binding sites, using custom Perl scripts.

#### Binding sites of transcription factors

We used Transfac Professional database ver.11.4 (Matys et al., 2006) and its associated Match program to map all binding sites

**Table 3 | Example of contingency table for mutation type overrepresentation test.**

Number of insertions altering any TFBS	Number of all mutations, excluding insertions, altering any TFBS
Number of insertions not altering TFBS	Number of all mutations, excluding insertions, not altering any TFBS

of vertebrate TFs to the promoter region. We used high quality matrices and optimized threshold setting for “minimum FP.” This allows for the TFBS predictions with presumed minimal number of false positive predictions.

### ENRICHMENT ANALYSIS

For the TFs that correspond to the above mentioned 25 TFBS types, we used GeneMANIA (Warde-Farley et al., 2010) program to find out potentially enriched GO categories associated with these TFs.

### STATISTICAL SIGNIFICANCE

To find overrepresented types of mutations within all TFBSs, we applied the right-sided exact Fisher’s test to contingency tables (example is shown in **Table 3**) with Bonferroni correction for multiplicity testing. For each mutation types, we calculated the total number of mutations of the considered type that overlapped any TFBS or fell outside of any predicted TFBS. As a background we used the total number of mutation of all other types that altered

**Table 4 | Example of contingency table for TF HNF3 alpha overrepresentation test.**

Number of HNF3 alpha binding sites altered by any mutations	Number of all, excluding HNF3 alpha binding sites, altered by any mutations
Number of HNF3 alpha binding sites not altered by mutations	Number of all, excluding HNF3 alpha binding sites, not altered by mutations

*The obtained p-values were corrected for multiplicity testing using Bonferroni method. The significance threshold of 0.05 was used. In the use of GeneMANIA we report the p-values generated by the tool and determine statistical significance using 0.05 as the threshold value.*

any TFBS or fell outside of any predicted TFBS. To find TFBS altered significantly by any particular type of mutations we applied the exact Fisher’s test to contingency tables (example is shown in the **Table 4**) with Bonferroni correction for multiplicity testing. For each TF, we calculated the total number of different mutations altering any TFBS for a given TF or falling outside of such TFBS. As a background we used the total number of TFBSs of all other TF containing or not containing any mutations.

### SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: [http://www.frontiersin.org/Statistical\\_Genetics\\_and\\_Methodology/10.3389/fgene.2012.00100/abstract](http://www.frontiersin.org/Statistical_Genetics_and_Methodology/10.3389/fgene.2012.00100/abstract)

### REFERENCES

- Alexiou, P., Vergoulis, T., Gleditsch, M., Prekas, G., Dalamagas, T., Megraw, M., Grosse, I., Sellis, T., and Hatzigeorgiou, A. G. (2010). miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic Acids Res.* 38, D137–D141.
- Altshuler, D. L., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Collins, F. S., De La Vega, F. M., Donnelly, P., Egholm, M., Flicek, P., Gabriel, S. B., Gibbs, R. A., Knoppers, B. M., Lander, E. S., LeFrach, H., Mardis, E. R., McVean, G. A., Nickerson, D. A., Peltonen, L., Schafer, A. J., Sherry, S. T., Wang, J., Wilson, R. K., Deiros, D., Metzker, M., Muzny, D., Reid, J., Wheeler, D., Wang, S. J., Li, J., Jian, M., Li, G., Li, R., Liang, H., Tian, G., Wang, B., Wang, J., Wang, W., Yang, H., Zhang, X., Zheng, H., Ambrogio, L., Bloom, T., Cibulskis, K., Fennell, T. J., Jaffe, D. B., Shefler, E., Sougnez, C. L., Bentley, I. D. R., Gormley, N., Humphray, S., Kingsbury, Z., Koko-Gonzales, P., Stone, J., Mc Kernan, K. J., Costa, G. L., Ichikawa, J. K., Lee, C. C., Sudbrak, R., Borodina, T. A., Dahl, A., Davydov, A. N., Marquardt, P., Mertes, E., Nietfeld, W., Rosenstiel, P., Schreiber, S., Soldatov, A. V., Timmermann, B., Tolzmann, M., Affourtit, J., Ashworth, D., Attiya, S., Bachorski, M., Buglione, E., Burke, A., Caprio, A., Celone, C., Clark, S., Conners, D., Desany, B., Gu, L., Guccione, L., Kao, K., Keibel, A., Knowlton, J., Labrecque, M., McDade, L., Mealmaker, C., Minderman, M., Nawrocki, A., Niazi, F., Pareja, K., Ramenani, R., Riches, D., Song, W., Turcotte, C., Wang, S., Doelling, D., Fulton, L., Fulton, R., Weinstock, G., Burton, J., Carter, D. M., Churcher, C., Coffey, A., Cox, A., Palotie, A., Quail, M., Skelly, T., Stalker, J., Swerdlow, H. P., Turner, D., De Witte, A., Giles, S., Bainbridge, M., Challis, D., Sabo, A., Yu, F., Yu, J., Fang, X., Guo, X., Li, Y., Luo, R., Tai, S., Wu, H., Zheng, H., Zheng, X., Zhou, Y., Marth, G. T., Garrison, E. P., Huang, W., Indap, A., Kural, D., Lee, W. P., Leong, W. F., Quinlan, A. R., Stewart, C., Stromberg, M. P., Ward, A. N., Wu, J., Lee, C., Mills, R. E., Shi, X., Daly, M. J., DePristo, M. A., Ball, A. D., Banks, E., Browning, B. L., Garimella, K. V., Grossman, S. R., Handsaker, R. E., Hanna, M., Hartl, C., Kernysky, A. M., Korn, J. M., Li, H., Maguire, J. R., McKenna, A., Nemesh, J. C., Philippakis, A. A., Poplin, R. E., Price, A., Rivas, M. A., Sabeti, P. C., Schaffner, S. F., Shlyakhter, I. A., Cooper, D. N., Ball, E. V., Mort, M., Phillips, A. D., Stenson, P. D., Sebat, J., Makarov, V., Ye, K., Yoon, S. C., Bustamante, C. D., Boyko, A., Degenhardt, J., Gravel, S., Gutenkunst, R. N., Kaganovich, M., Keinan, A., Lacroite, P., Ma, X., Reynolds, A., Clarke, L., Cunningham, F., Herrero, J., Keenen, S., Kulesha, E., Leinonen, R., McLaren, W. M., Radhakrishnan, R., Smith, R. E., Zalunin, V., Korbel, J. O., Stütz, A. M., Humphray, I. S., Bauer, M., Cheetham, R. K., Cox, T., Eberle, M., James, T., Kahn, S., Murray, L., Ye, K., Fu, Y., Hyland, F. C. L., Manning, J. M., Stephen, F., McLaughlin, P. H. E., Sakarya, O., Sun, Y. A., Tsung, E. F., Mark, A., Batzer, K. M. K., Walker, J. A., Albrecht, M. W., Amstislavskiy, V. S., Herwig, R., Parkhomchuk, D. V., Agarwala, R., Khouri, H. M., Morgulis, A. O., Paschall, J. E., Phan, L. D., Rotmistrovsky, K. E., Sanders, R. D., Shumway, M. F., Xiao, C., Gil, A., McVean, A. A., Iqbal, Z., Lunter, G., Marchini, J. L., Moutsianas, L., Myers, S., Tumian, A., Knight, J., Winer, R., Craig, D. W., Beckstrom-Sternberg, S. M., Christoforides, A., Kurdoglu, A. A., Pearson, J. V., Sinari, S. A., Tembe, W. D., Haussler, D., Hinrichs, A. S., Katzman, S. J., Kern, A., Kuhn, R. M., Przeworski, M., Hernandez, R. D., Howie, B., Kelley, J. L., Melton, S. C., Li, Y., Anderson, P., Blackwell, T., Chen, W., Cookson, W. O., Ding, J., Kang, H. M., Lathrop, M., Liang, L., Moffatt, M. F., Scheet, P., Sidore, C., Snyder, M., Zhan, X., Zöllner, S., Awadalla, P., Casals, F., Idaghdour, Y., Keebler, J., Stone, E. A., Zilvermit, M., Jorde, L., Xing, J., Eichler, E. E., Aksay, G., Alkan, C., Hajirasouliha, I., Hormozdiari, F., Kidd, J. M., CenkSahinalp, S., Sudmant, P. H., Chen, K., Chinwalla, A., Ding, L., Koboldt, D. C., McLellan, M. D., Wallis, J. W., Wendl, M. C., Zhang, Q., Albers, C. A., Ayub, Q., Balasubramaniam, S., Barrett, J. C., Chen, Y., Conrad, D. F., Daneczek, P., Dermitzakis, E. T., Hu, M., Huang, N., Hurles, M. E., Jin, H., Jostins, L., Keane, T. M., Quang Le, S., Lindsay, S., Long, Q., MacArthur, D. G., Montgomery, S. B., Parts, L., Tyler-Smith, C., Walter, K., Zhang, Y., Gerstein, M. B., Snyder, M., Abyzov, A., Balasubramanian, S., Bjornson, R., Grubert, F., Habegger, L., Haraksingh, R., Khurana, E., Lam, H. Y. K., Leng, J., Mu, X. J., Urban, A. E., Zhang, Z., McCarroll, S. A., Zheng-Bradley, X., Batzer, M. A., Hurles, M. E., Du, J., Jee, J., Coafra, C., Dinh, H., Kovar, C., Lee, S., Nazareth, L., Wilkinson, J., Coffey, A., Scott, C., Tyler-Smith, C., Gharani, N., Kaye, J. S., Kent, A., Li, T., McGuire, A. L., Ossorio, P. N., Rotimi, C. N., Su, Y., Toji, L. H., Felsenfeld, A. L., McEwen, J. E., Abdallah, A., Juenger, C. R.,

- Clemm, N. C., Duncanson, A., Green, E. D., Guyer, M. S., and Peterson, J. L. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073.
- Amberger, J., Bocchini, C. A., Scott, A. F., and Hamosh, A. (2009). McKusick's online mendelian inheritance in man (OMIM). *Nucleic Acids Res.* 37, D793–D796.
- Antigny, F., Norez, C., Becq, F., and Vandebrouck, C. (2011). CFTR and Ca signaling in cystic fibrosis. *Front. Pharmacol.* 2:67. doi:10.3389/fphar.2011.00067
- Bao, L., Zhou, M., Wu, L., Lu, L., Goldowitz, D., Williams, R. W., and Cui, Y. (2007). PolymiRTS database: linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids Res.* 35, D51–D54.
- Batist, G., Wu, J. H., Spatz, A., Miller, W. H., Cocolakis, E., Rousseau, C., Diaz, Z., Ferrario, C., and Basik, M. (2011). Resistance to cancer treatment: the role of somatic genetic events and the challenges for targeted therapies. *Front. Pharmacol.* 2:59. doi:10.3389/fphar.2011.00059
- Belmont, J. W., Boudreau, A., Leal, S. M., Hardenbol, P., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Parkin, M., Roy, J., Stahl, E., Winchester, E., Altschuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Wayne, M. M. Y., Tsui, S. K. W., Xue, H., Wong, J. T. F., Galver, L. M., Fan, J. B., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallée, C., Verner, A., Hudson, T. J., Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Kwok, P.-Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.-C., Mak, W., Sham, P. C., Song, Y. Q., Tam, P. K. H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stanger, B. E., Whittaker, P., Bentley, D. R., De Bakker, P. I. W., Barrett, J., Fry, B., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Varilly, P., Stein, L. D., Krishnan, L., Smith, A. V., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Botto, L., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Tsunoda, T., Mullikin, J. C., Sherry, S. T., Feolo, M., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adewole, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimma, C., Royal, C. D. M., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Gibbs, R. A., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Yakub, I., Gabriel, S. B., Onofrio, R. C., Ziaugra, L., Birren, B. W., Daly, M. J., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Bird, C. R., Guyer, M. S., Nailer, P. J., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Witonsky, J., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R., and Stewart, J. (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.
- Bennett, M. J. (2010). Galactosemia diagnosis gets an upgrade. *Clin. Chem.* 56, 690–692.
- Callahan, J. W., and Abercrombie, E. D. (2011). In vivo dopamine efflux is decreased in striatum of both fragment (R6/2) and full-length (YAC128) transgenic mouse models of Huntington's disease. *Front. Syst. Neurosci.* 5:61. doi:10.3389/fnsys.2011.00061
- Collombat, P., Xu, X., Ravassard, P., Sosa-Pineda, B., Dussaud, S., Billestrup, N., Madsen, O. D., Serup, P., Heimberg, H., and Mansouri, A. (2009). The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into alpha and subsequently beta cells. *Cell* 138, 449–462.
- Conde, L., Vaquerizas, J. M., Dopazo, H., Arbiza, L., Reumers, J., Rousseau, F., Schymkowitz, J., and Dopazo, J. (2006). PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res.* 34, W621–W625.
- Decker, T., Pasca di Magliano, M., McManus, S., Sun, Q., Bonifer, C., Tagoh, H., and Busslinger, M. (2009). Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis. *Immunity* 30, 508–520.
- Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., Goldman, M., Barber, G. P., Clawson, H., Coelho, A., Diekhans, M., Dreszer, T. R., Giardine, B. M., Harte, R. A., Hillman-Jackson, J., Hsu, F., Kirkup, V., Kuhn, R. M., Learned, K., Li, C. H., Meyer, L. R., Pohl, A., Raney, B. J., Rosenbloom, K. R., Smith, K. E., Haussler, D., and Kent, W. J. (2011). The UCSC genome browser database: update 2011. *Nucleic Acids Res.* 39, D876–D882.
- González-Hernández, T., Cruz-Muros, I., Afonso-Oramas, D., Salas-Hernandez, J., and Castro-Hernandez, J. (2010). Vulnerability of mesostriatal dopaminergic neurons in Parkinson's disease. *Front. Neuroanat.* 4:140. doi:10.3389/fnana.2010.00140
- Gu, Y., Harley, I. T. W., Henderson, L. B., Aronow, B. J., Vietor, I., Huber, L. A., Harley, J. B., Kilpatrick, J. R., Langefeld, C. D., Williams, A. H., Jegga, A. G., Chen, J., Wills-Karp, M., Arshad, S. H., Ewart, S. L., Thio, C. L., Flick, L. M., Filippi, M. D., Grimes, H. L., Drumm, M. L., Cutting, G. R., Knowles, M. R., and Karp, C. L. (2009). Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease. *Nature* 458, 1039–1042.
- Guo, Z., Packard, A., Krolewski, R. C., Harris, M. T., Manglapus, G. L., and Schwob, J. E. (2010). Expression of pax6 and sox2 in adult olfactory epithelium. *J. Comp. Neurol.* 518, 4395–4418.
- Hamacher, R., Diersch, S., Scheibel, M., Eckel, F., Mayr, M., Rad, R., Bajbouj, M., Schmid, R. M., Saur, D., and Schneider, G. (2009). Interleukin 1 beta gene promoter SNPs are associated with risk of pancreatic cancer. *Cytokine* 46, 182–186.
- Hannenhalli, S., and Kaestner, K. H. (2009). The evolution of Fox genes and their role in development and disease. *Nat. Rev. Genet.* 10, 233–240.
- Hariharan, M., Scaria, V., and Brahmachari, S. K. (2009). dbSMR: a novel resource of genome-wide SNPs affecting microRNA mediated regulation. *BMC Bioinformatics* 10, 108. doi:10.1186/1471-2105-10-108
- Heckmann, J. M., Uwimpuhwe, H., Ballo, R., Kaur, M., Bajic, V. B., and Prince, S. (2010). A functional SNP in the regulatory region of the decay-accelerating factor gene associates with extraocular muscle paresis in myasthenia gravis. *Genes Immun.* 11, 1–10.
- Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B. E., Martin, M. J., McGarvey, P., and Gasteiger, E. (2009). Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* 10, 136. doi:10.1186/1471-2105-10-136
- Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S. M., Habetger, L., Rozowsky, J., Shi, M., Urban, A. E., Hong, M. Y., Karczewski, K. J., Huber, W., Weissman, S. M., Gerstein, M. B., Korbel, J. O., and Snyder, M. (2010). Variation in transcription factor binding among humans. *Science* 328, 232–235.
- Kim, B. C., Kim, W. Y., Park, D., Chung, W. H., Shin, K. S., and Bhak, J. (2008). SNP@Promoter: a database of human SNPs (single nucleotide polymorphisms) within the putative promoter regions. *BMC Bioinformatics* 9(Suppl. 1), S2. doi:10.1186/1471-2105-9-S1-S2
- Kopp, M., and Hermisson, J. (2009). The genetic basis of phenotypic adaptation I: fixation of beneficial mutations in the moving optimum model. *Genetics* 182, 233–249.
- Kumar, P., Henikoff, S., and Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1081.
- Lim, D. H. K., Rehal, P. K., Nahorski, M. S., Macdonald, F., Claessens, T., Van Geel, M., Gijzen, L., Gille, J. J. P., Giraud, S., Richard, S., van Steensel, M., Menko, F. H., and Maher, E. R. (2010). A new locus-specific database (LSDB) for mutations in the folliculin (FLCN) gene. *Hum. Mutat.* 31, E1043–E1051.
- Lindoso, R. S., Verdoorn, K. S., and Eincker-Lamas, M. (2009). Renal recovery after injury: the role of Pax-2. *Nephrol. Dial. Transplant.* 24, 2628–2633.
- Ma, W., Kantarjian, H., Zhang, K., Zhang, X., Wang, X., Chen, C., Donahue, A. C., Zhang, Z., Yeh, C.-H., O'Brien, S., Garcia-Manero, G., Caporaso, N., Landgren, O.,

- and Albitar, M. (2010). Significant association between polymorphism of the erythropoietin gene promoter and myelodysplastic syndrome. *BMC Med. Genet.* 11, 163. doi:10.1186/1471-2350-11-163
- Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, a, Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A. E., and Wingender, E. (2006). TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* 34, D108–D110.
- Pruitt, K. D., Tatusova, T., Klimke, W., and Maglott, D. R. (2009). NCBI reference sequences: current status, policy and new initiatives. *Nucleic Acids Res.* 37, D32–D36.
- Rowan, S., Siggers, T., Lachke, S. A., Yue, Y., Bulyk, M. L., and Maas, R. L. (2010). Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes Dev.* 24, 980–985.
- Roze, E. (2011). Huntington's disease and striatal signaling. *Front. Neuroanat.* 5:55. doi:10.3389/fnana.2011.00055
- Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., Dicuccio, M., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Lu, Z., Madden, T. L., Madej, T., Maglott, D. R., Marchler-Bauer, A., Miller, V., Mizrachi, I., Ostell, J., Panchenko, A., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T. A., Wagner, L., Wang, Y., John Wilbur, W., Yaschenko, E., and Ye, J. (2010). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 38, D5–D16.
- Schmeier, S., Schaefer, U., MacPherson, C. R., and Bajic, V. B. (2011). dPORE-miRNA: polymorphic regulation of microRNA genes. *PLoS ONE* 6, e21887. doi:10.1371/journal.pone.0016657
- Song, F., Li, X., Zhang, M., Yao, P., Yang, N., Sun, X., Hu, F. B., and Liu, L. (2009). Association between heme oxygenase-1 gene promoter polymorphisms and type 2 diabetes in a Chinese population. *Am. J. Epidemiol.* 170, 747–756.
- Stenson, P. D., Mort, M., Ball, E. V., Howells, K., Phillips, A. D., Thomas, N. S., and Cooper, D. N. (2009). The human gene mutation database: 2008 update. *Genome Med.* 1, 13.
- Wardle-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., and Morris, Q. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38, W214–W220.
- Yu, Y., Keller, S. H., Remillard, C. V., Safrina, O., Nicholson, A., Zhang, S. L., Jiang, W., Vangala, N., Landsberg, J. W., Wang, J.-Y., Thistlethwaite, P. A., Channick, R. N., Robbins, I. M., Loyd, J. E., Ghofrani, H. A., Grimminger, F., Schermuly, R. T., Cahalan, M. D., Rubin, L. J., and Yuan, J. X. (2009). A functional single-nucleotide polymorphism in the TRPC6 gene promoter associated with idiopathic pulmonary arterial hypertension. *Circulation* 119, 2313–2322.
- Zhang, H., Chen, H., Luo, H., An, J., Sun, L., Mei, L., He, C., Jiang, L., Jiang, W., Xia, K., Li, J. D., and Feng, Y. (2012). Functional analysis of Waardenburg syndrome-associated PAX3 and SOX10 mutations: report of a dominant-negative SOX10 mutation in Waardenburg syndrome type II. *Hum. Genet.* 131, 491–503.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 November 2011; accepted: 16 May 2012; published online: 01 June 2012.

Citation: Kamanu FK, Medvedeva YA, Schaefer U, Jankovic BR, Archer JAC and Bajic VB (2012) Mutations and binding sites of human transcription factors. *Front. Genet.* 3:100. doi: 10.3389/fgene.2012.00100

This article was submitted to *Frontiers in Statistical Genetics and Methodology*, a specialty of *Frontiers in Genetics*.

Copyright © 2012 Kamanu, Medvedeva, Schaefer, Jankovic, Archer and Bajic. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.