Electronic supplementary information

Structural and conformational changes induced by missense variants in the zinc finger domains of GATA3 involved in breast cancer

Rakesh Kumar[1], Rahul Kumar[1] Pranay Tanwar[1]*, SVS Deo[1], Sandeep Mathur[2], Usha Agarwal[3], and Showket Hussain[4]

[1]Dr.B.R.A.-Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, INDIA-110029

[2]Department of Pathology, All India Institute of Medical Sciences, New Delhi, INDIA-110029

[3]National Institute of Pathology, New Delhi, INDIA-110029

[4]Division of Molecular Oncology, National Institute of Cancer Prevention and Research, Noida, INDIA-201301

*__Corresponding author:__ Dr. Pranay Tanwar

Dr.B.R.A.-Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, INDIA-110029

E-mail: pranaytanwar@gmail.com

Supplementary figures	: 11
Supplementary tables	: 7

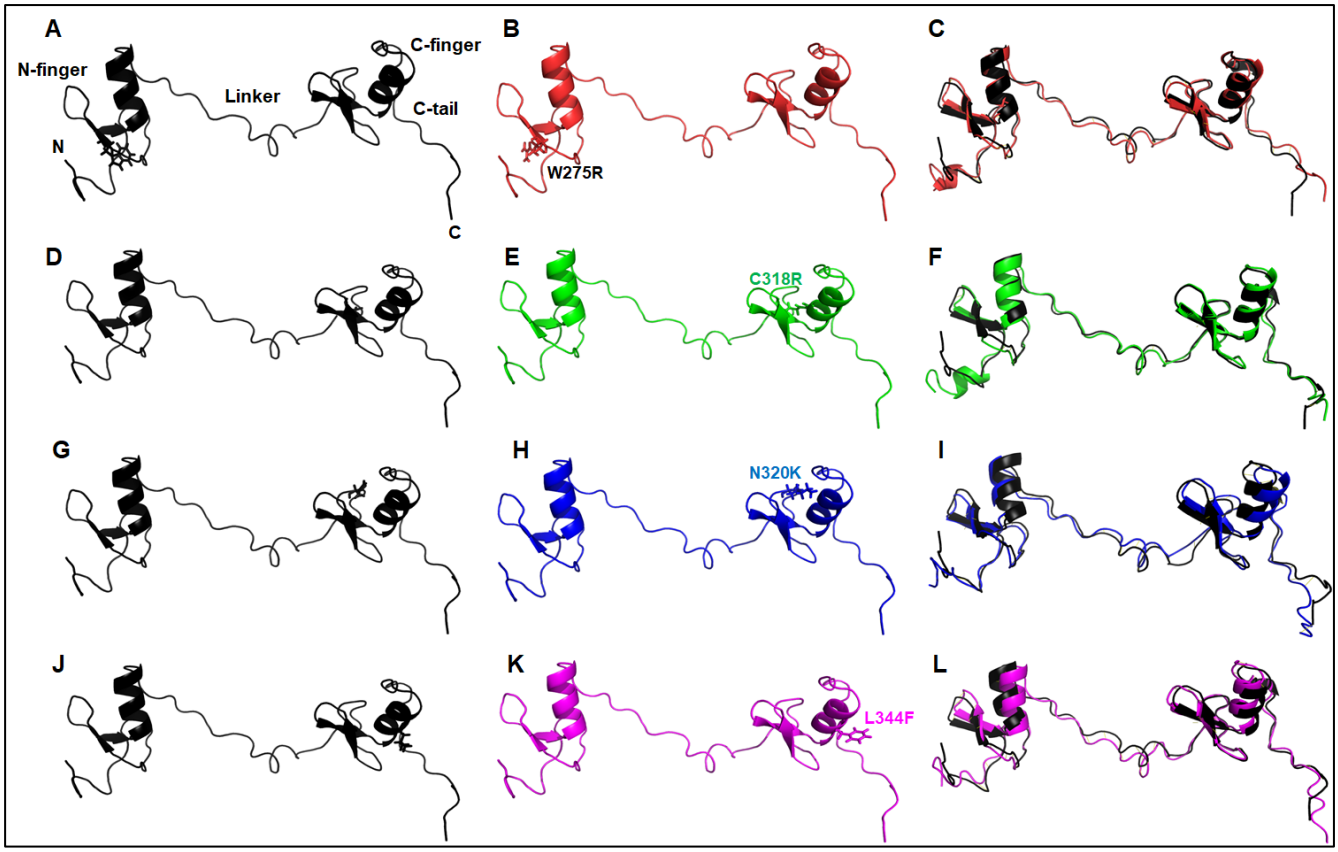

Fig. S1 **RMSD calculation between WT and MT proteins.** (A-C) WT, W275R and superimpose both the protein, (D-F) WT, C318R and superimpose both the protein, (G-I) WT, N320K and superimpose both the protein and (J-L) WT, L344F and superimpose both the protein. All 3D models and mutant residues were depicted in uniform cartoon colour (WT: black, W275R: red, C318R: green, N320K: blue, L344F: magenta) and stick modes, respectively.

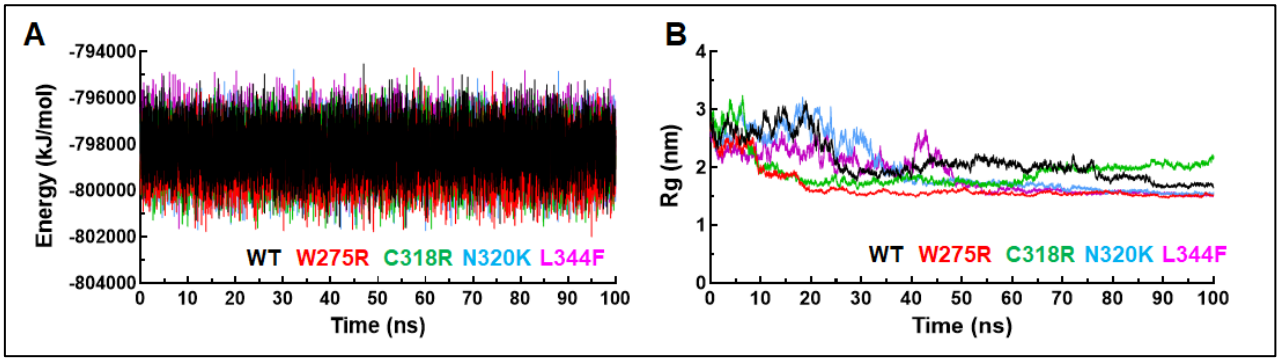

Fig. S2 **MD simulation.** (A) Potential energies and (B) Radius of gyration of backbone proteins of WT and all MTs. WT, W275R, C318R, N320K and L344F MTs were labelled in black, red, green, blue and magenta lines in 2D graphs, respectively.

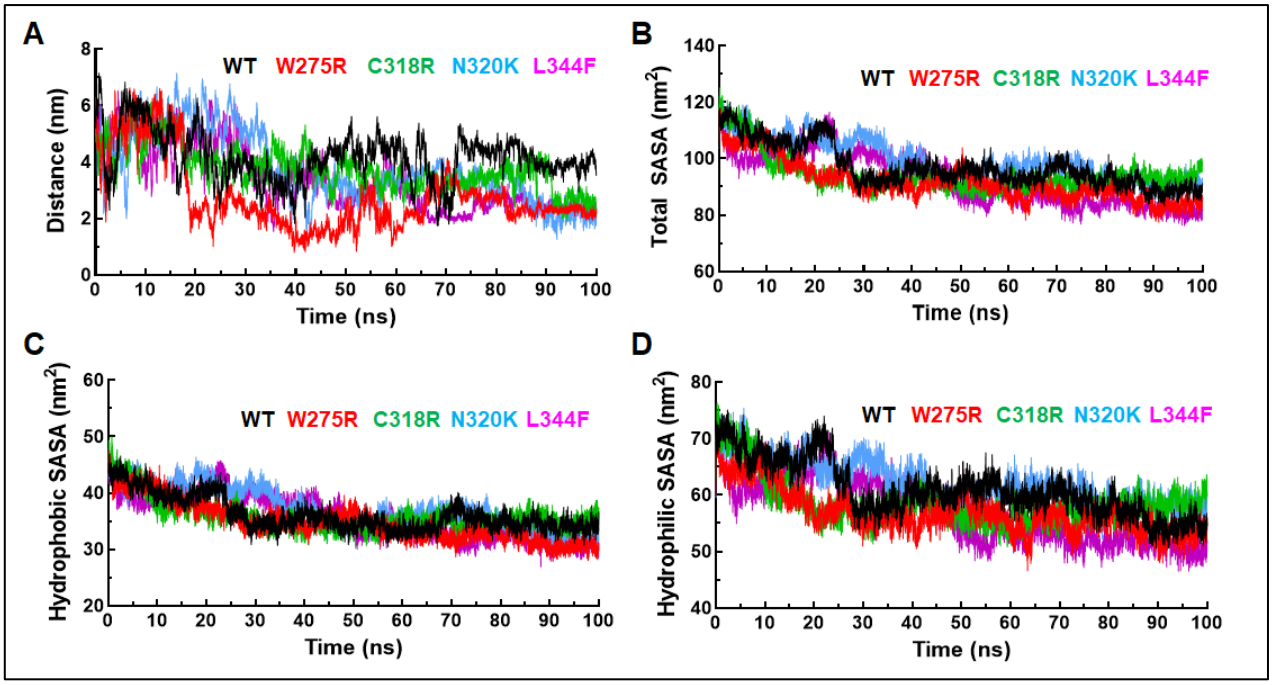

Fig. S3 **Distance and SASA analyses. (A)** Distance between N- and C-terminals of WT and MT proteins, **(B)** Total SASA, **(C)** Hydrophobic SASA and **(D)** Hydrophilic SASA of WT and all MTs. MTs were labelled in black, red, green, blue and magenta lines in 2D graphs, respectively.

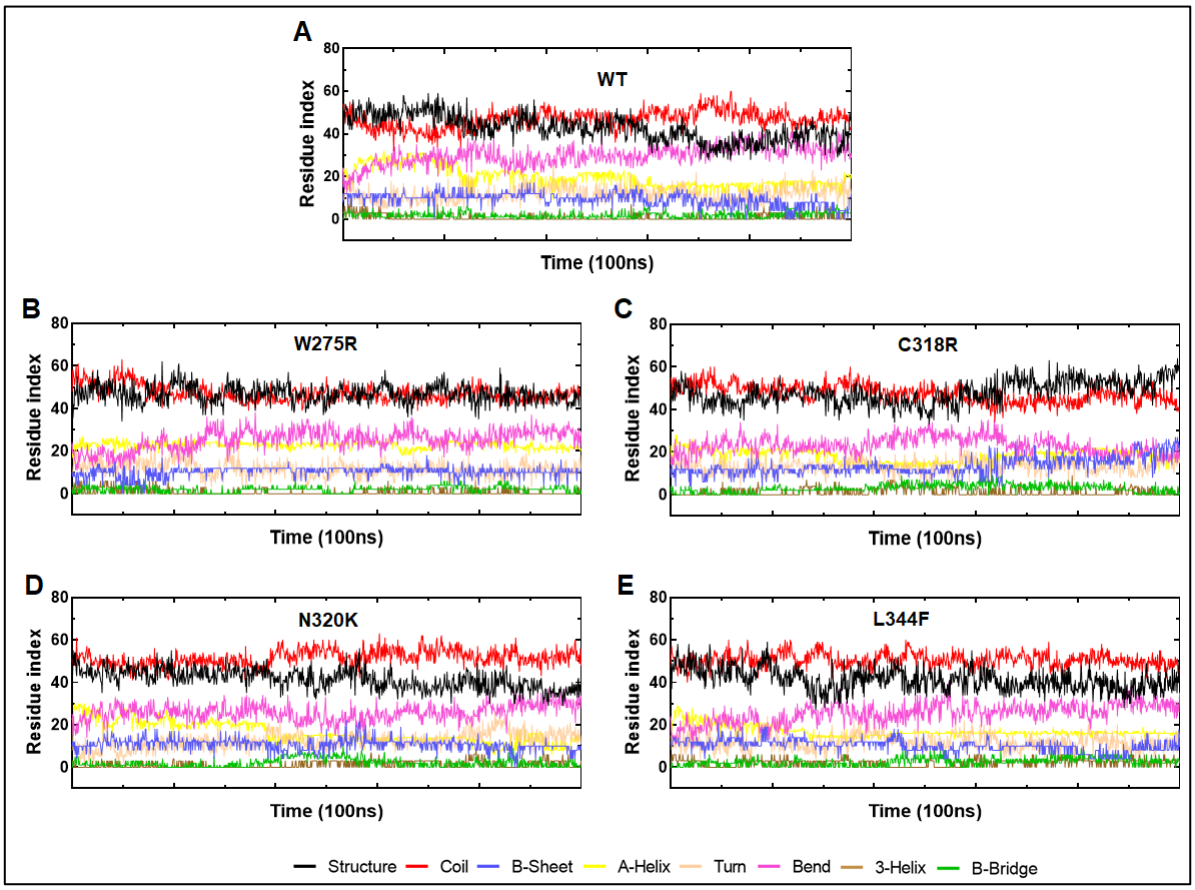

Fig. S4 **Secondary structures analyses of (A)** WT, **(B)** W275R, **(C)** C318R **(D)** N320K and **(E)** L344F MTs. Different secondary structures moieties such as structure, coil, β-sheet, α-helix, turn, bend,

3-helix and β-bridge were labelled as black, red, blue, yellow, magenta, brown and green colour, respectively.

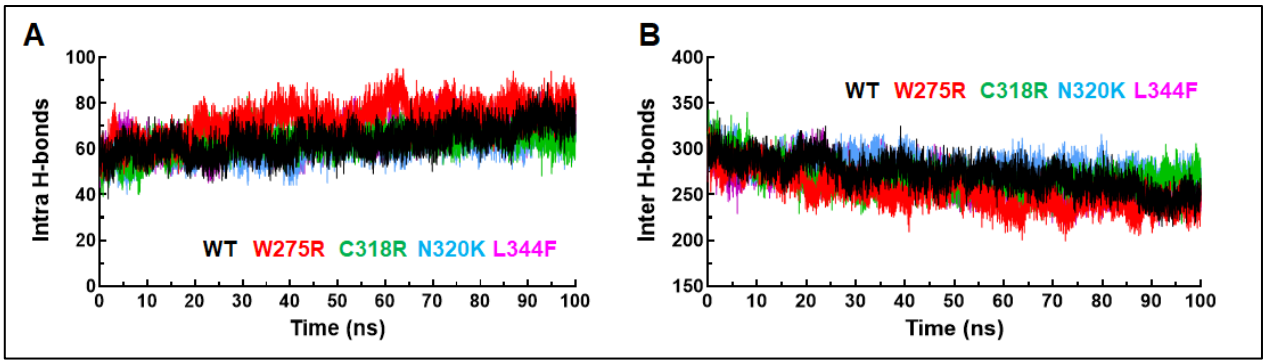

Fig. S5 Hydrogen bond analyses of WT and MTs GATA3 (A) Intra or protein-protein H-bonds and **(B)** Inter or protein-solvent H-bonds.

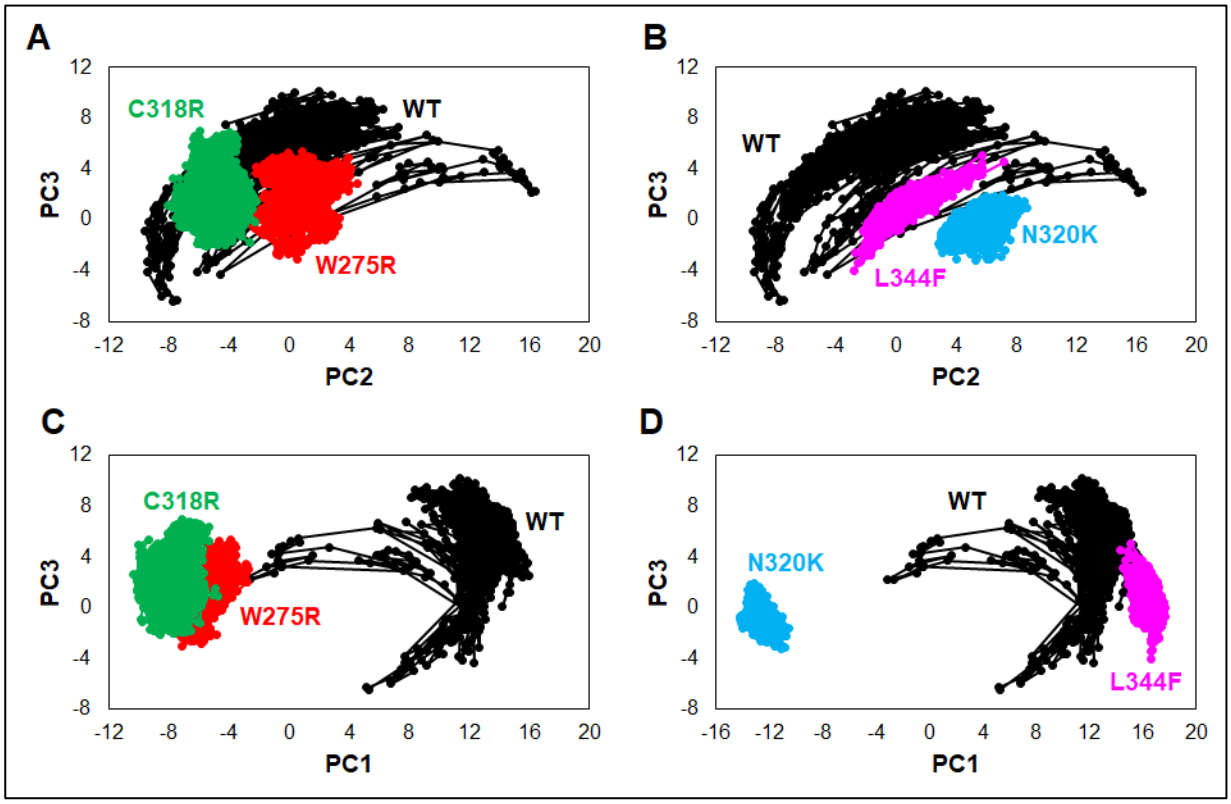

Fig. S6 Collective mode of motions and essential dynamics analyses. (A) and **(B)** Projection of principle component 2 and 3 of WT and all MTs. **(C)** and **(D)** Projection of principle component 1 and 3 of WT and all MTs.

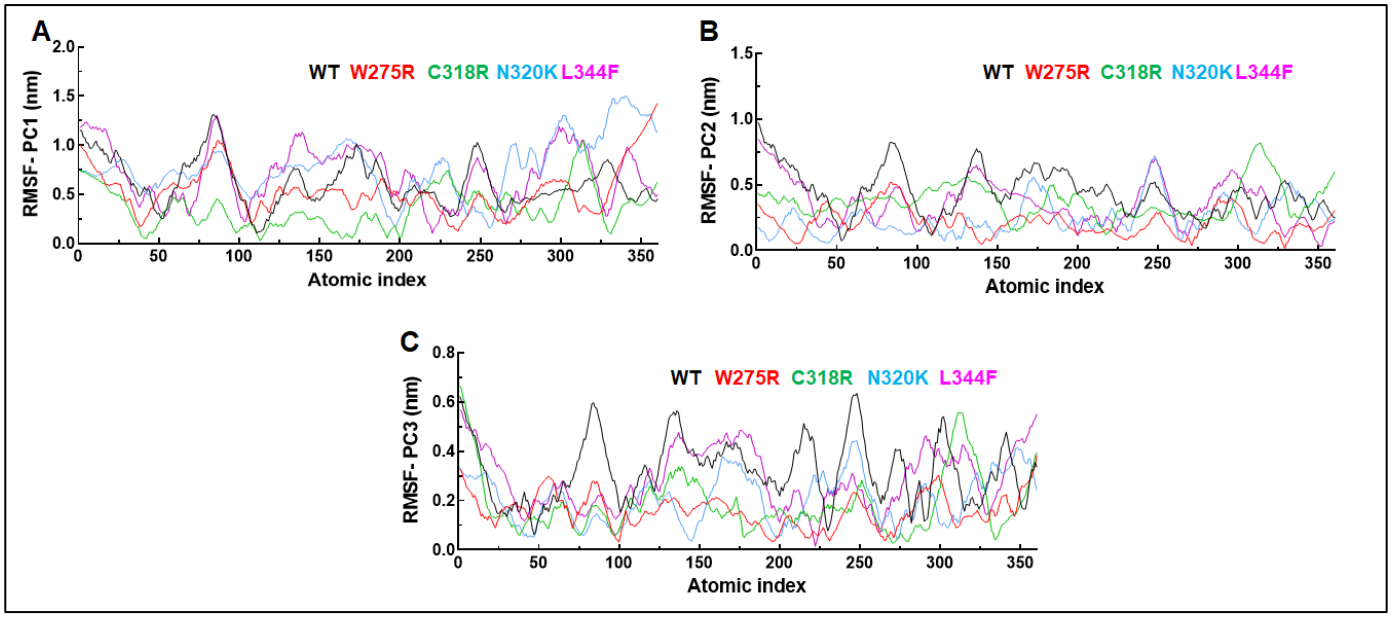

Fig. S7 **Root mean square fluctuation analyses of (A)** Principle component 1 **(B)** Principle component 2 and **(C)** Principle component 3 of WT and all MTs.

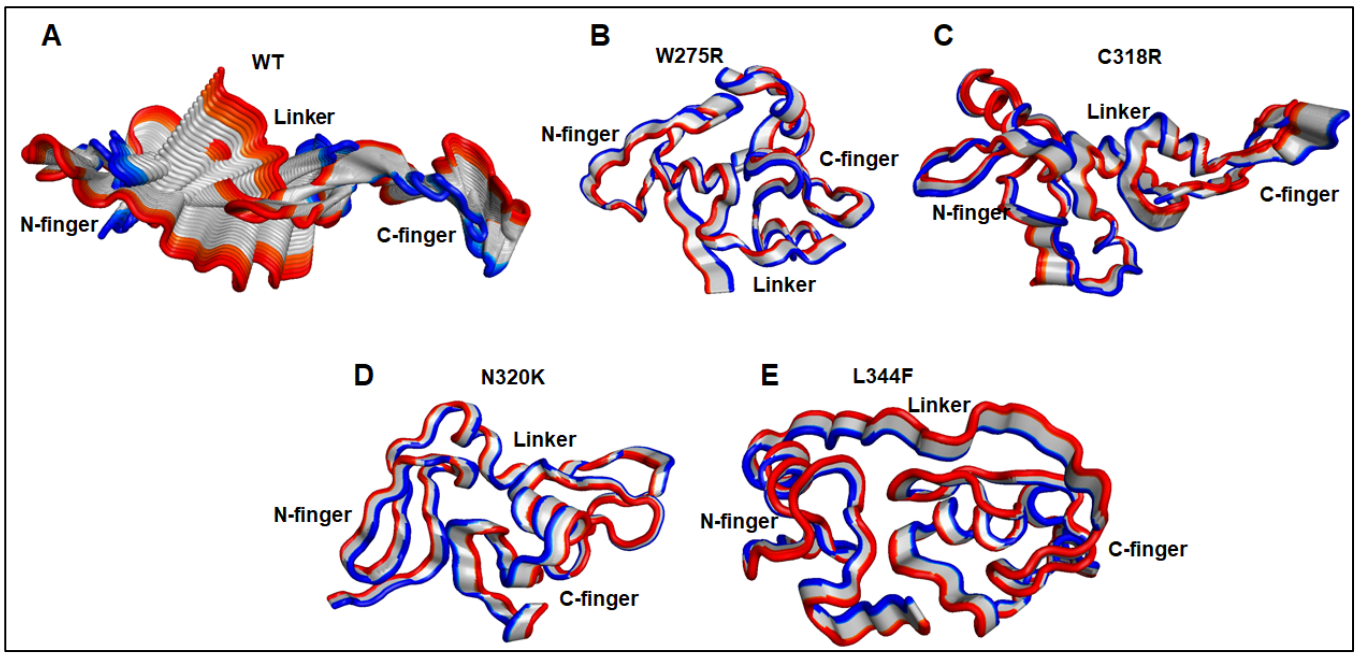

Fig. S8 **Sequentially superimpositions of 30 frames from first PC of (A)** WT, **(B)** W275R, **(C)** C318R **(D)** N320K and **(E)** L344F MTs. Red and blue represented minimum to maximum motions.

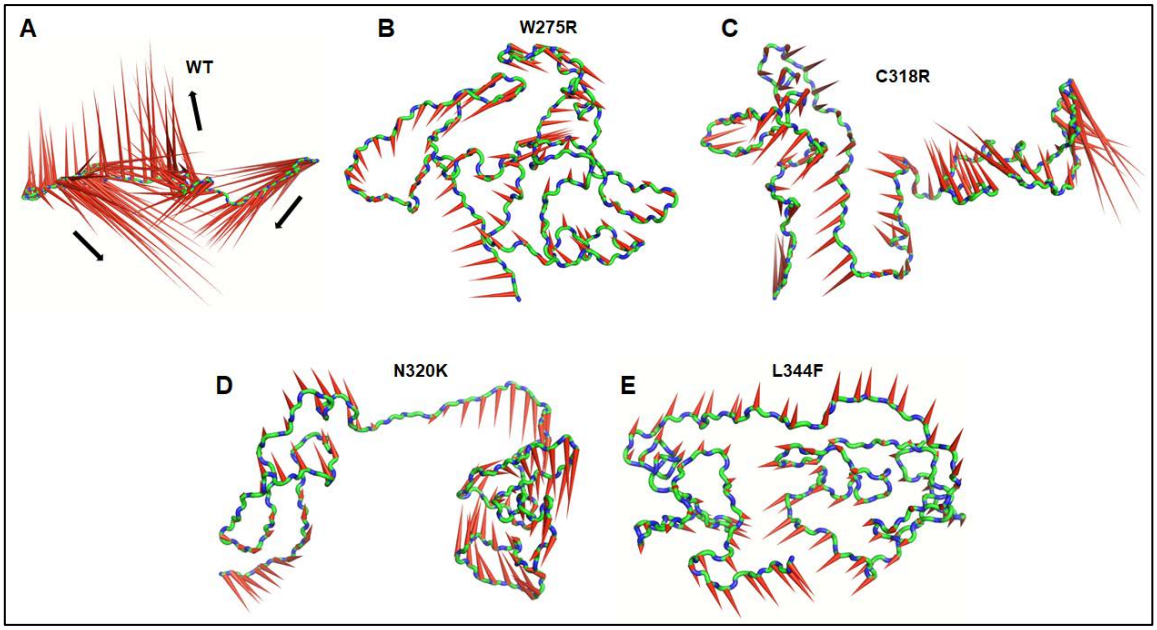

Fig. S9 **Porcupine structures derived from first PC of (A)** WT, **(B)** W275R, **(C)** C318R **(D)** N320K and **(E)** L344F MTs. Arrows and length of cones represented direction and magnitude of motions. Black solid arrow indicated direction of motions.

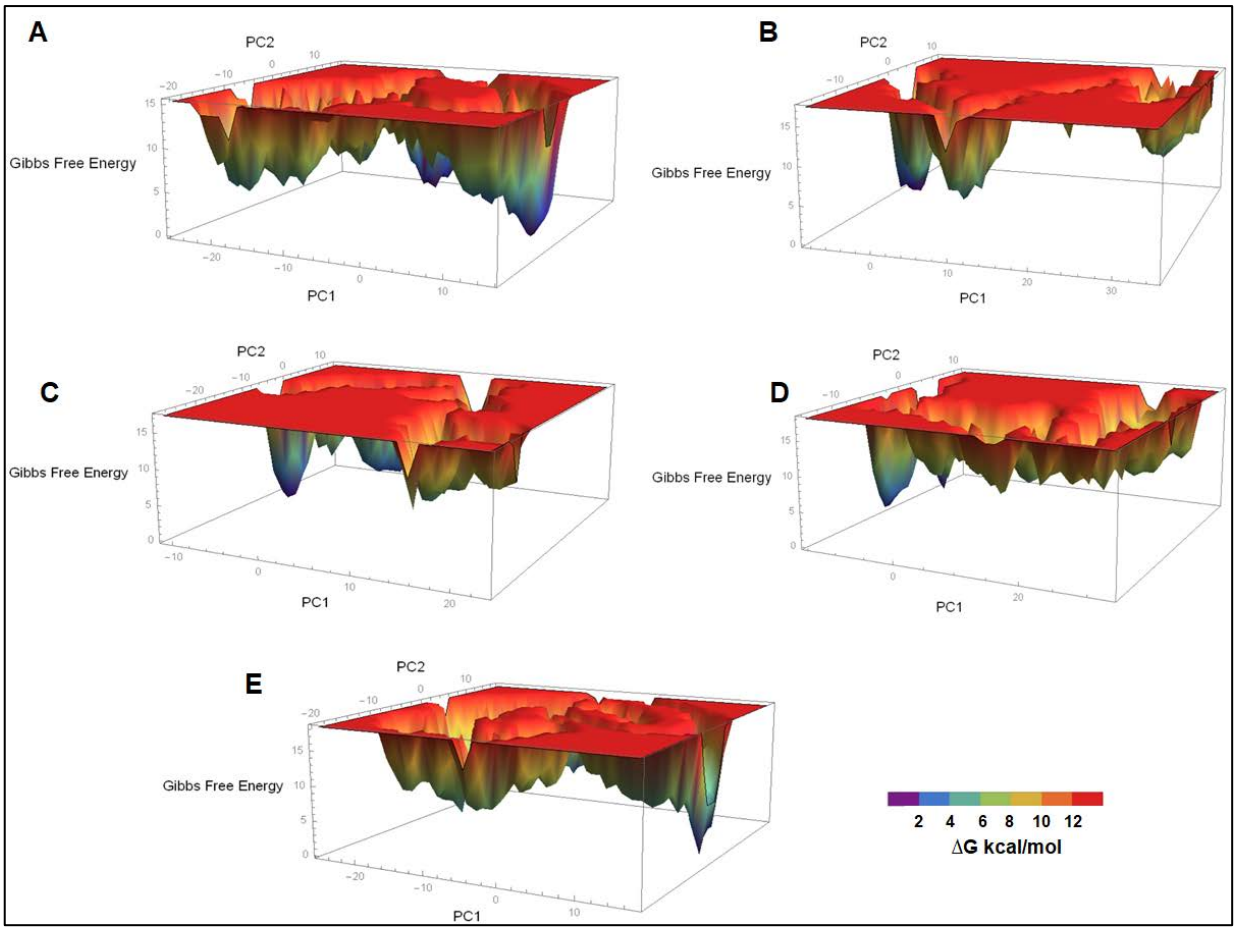

Fig. S10 **3D free energy landscape plots of (A)** WT, **(B)** W275R, **(C)** C318R, **(D)** N320K and **(E)** L344F MTs. ΔG were measured in kilocalorie per mol.

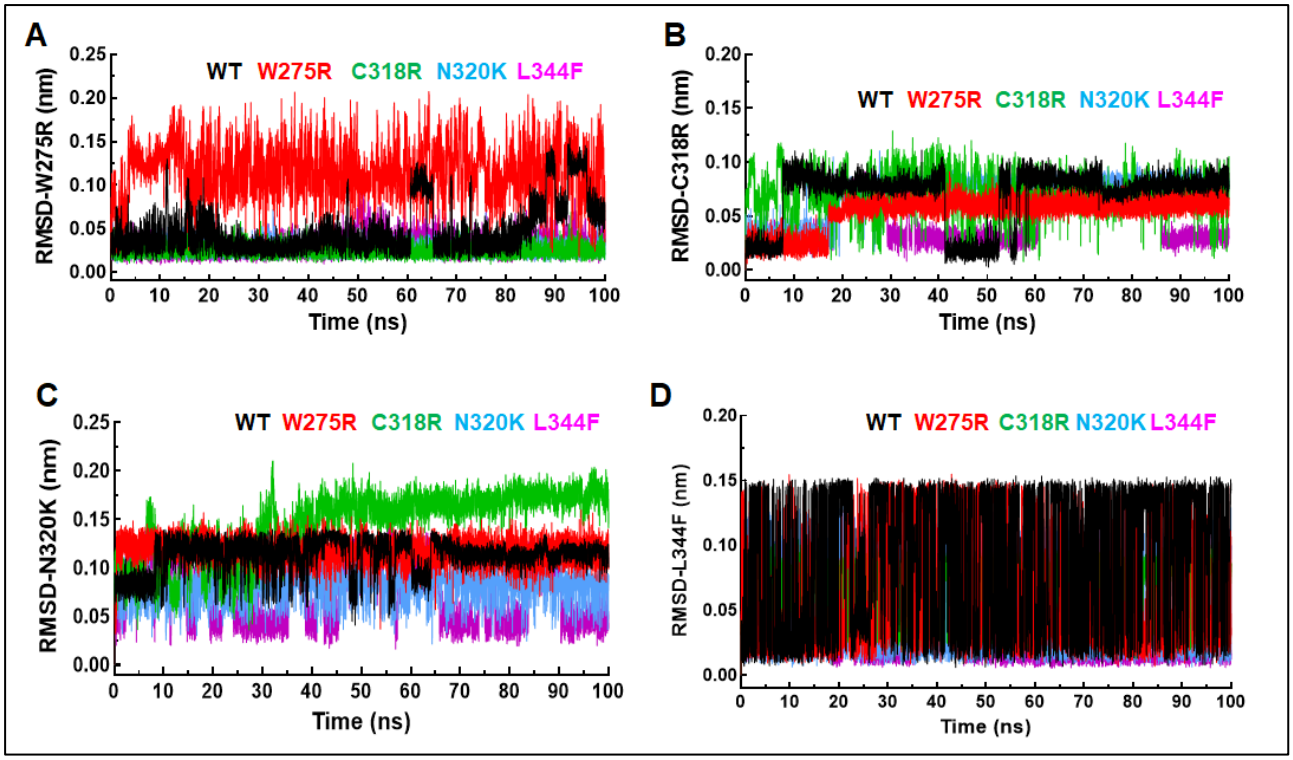

Fig. S11 RMSD of WT and MTs in different mutant's background of A) W275R, **(B)** C318R, **(C)** N320K and **(D)** L344F MTs.

Table S1. Prediction of functional consequences of human GATA3 missense variants.

Variants	Predict SNP	MAPP	PhD-SNP	Polyphen	SIFT	SNAP	
W275R	87% Deleterious	88% Deleterious	88% Deleterious	74% Deleterious	79% Deleterious	89% Deleterious	Disease
C318R	76% Deleterious	88% Deleterious	86% Deleterious	67% Deleterious	79% Deleterious	89% Deleterious	Disease
N320K	76% Deleterious	84% Deleterious	88% Deleterious	67% Deleterious	79% Deleterious	85% Deleterious	Disease
L344F	87% Deleterious	77% Deleterious	88% Deleterious	59% Deleterious	79% Deleterious	81% Deleterious	Disease

Table S2. Functional annotations of different missense variants of GATA3

Variants	PROVEAN	Mutation Assessor	FATHMM	PANTHER-PSEP
W275R	Deleterious	High impact	Cancer	Probably damaging
C318R	Deleterious	High impact	Cancer	Probably damaging
N320K	Deleterious	Medium impact	Cancer	Probably damaging
L344F	Deleterious	High impact	Cancer	Probably damaging

Table S3. Models validation and structures quality estimation

	Ramachandran plot			ERRAT	ProSA Z-Score	QMEAN Score	RMSD (Å)
	Favoured region	Allowed region	Disallowed region				
WT	68.60%	29.40%	2%	91.58%	-4.93	-5.62	
W275R	68.60%	29.40%	2%	80.10%	-4.92	-4.08	1.2
C318R	68.60%	29.40%	2%	75.45%	-4.59	-5.93	1.1
N320K	68.60%	29.40%	2%	90.90%	-4.9	-5.6	1.5
L344F	68.60%	29.40%	2%	91.58%	-4.74	-5.63	1.6

Table S4. Secondary structures percentage

	Coil	β-Sheet	B-Bridge	Bend	Turn	α-Helix
WT	39	8	2	25	10	16
W275R	39	9	2	21	10	19
C318R	40	11	2	19	11	15
N320K	43	9	2	22	10	14
L344F	42	8	2	21	10	14

Table S5. Cosine content analyses of first 3 PCs

		0-25 ns	25-50 ns	50-75 ns	75-100 ns
WT	**PC1**	0.6	0.24	0.36	0.004
	PC2	0.03	0.48	0.17	0.002
	PC3	0.63	0.24	0.24	0.004
W275R	**PC1**	0.34	0.05	0.007	0.0007
	PC2	0.46	0.58	0.0000025	0.1
	PC3	0.008	0.13	0.68	0.16
C318R	**PC1**	0.09	0.3	0.09	0.003
	PC2	0.8	0.02	0.85	0.00000054
	PC3	0.21	0.02	0.5	0.0008
N320K	**PC1**	0.001	0.87	0.004	0.0000074
	PC2	0.59	0.017	0.45	0.0004
	PC3	0.15	0.05	0.53	0.18
L344F	**PC1**	0.003	0.02	0.08	0.0003
	PC2	0.05	0.007	0.27	0.15
	PC3	0.016	0.02	0.03	0.06

Table S6. Cumulative percentages of first few PCs

	WT	W275R	C318R	N320K	L344F
Covariance matrix	386.56 nm^2	196.18 nm^2	178.54 nm^2	366.9 nm^2	400.9 nm^2
First 15 PCs	96.06%	95.73%	95.25%	96.84%	96.96%
First 10 PCs	93.26%	93.04%	92.02%	94.51%	94.52%
First 3 PCs	72.5%	78.91%	75.72%	81.95%	75.32%

Table S7. Residue-residue interactions in RIN

Positions	WT	W275R	C318R	N320K	L344F
275	R253, R262, T270, S271, T272, C285	**R262, Y283**	R253, R262, T270, S271, T272, C285	R253, R262, T270, S271, T272, C282	R253, R262, T270, S271, T272, C285
318	C318, N320, T323, V338, C339	N322, T323, V338, C339, C341, C342	**N320, T323, V338, C339, C341, C342**	C318, K322, T323, V338, C339	C318, N320, T323, V338, C339
320	V338, C342, Y348, N352	V338, C342, Y348, N352	V338, C342, Y348, N352	**V38, C342, Y346, N352, R353**	V338, C342, Y348, N352
344	N340, K347, L348	N340, K347, L348	N342, K347, L348	N340, K347, L348	**N90, K97, L98**